# Impact of Size Effects and Anomalous Skin Effect on Metallic Wires as GSI Interconnects

A Thesis Presented to
The Academic Faculty

by

Reza Sarvari

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
December 2008

# Impact of Size Effects and Anomalous Skin Effect on Metallic Wires as GSI Interconnects

Approved by:


Dr. James D. Meindl, Advisor
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Dr. Jeffrey A. Davis
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Dr. Thomas K. Gaylord
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Dr. Dennis W. Hess
School of Chemical and Biomolecular Engineering
*Georgia Institute of Technology*

Dr. Andrew F. Peterson
School of Electrical and Computer Engineering
*Georgia Institute of Technology*


Date Approved: August 19, 2008

*To my parents,*

*for their endless love and support.*

# ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor, Professor James D. Meindl, for being an excellent advisor, and for his support, inspiration, and kindness. It was a great honor and privilege to work under his tutelage. During my years at Georgia Tech, I have learned invaluable lessons from him not only about GSI and Interconnect, but also about thinking and learning, that I will carry throughout my life.

I would like to thank my committee members, Dr. Thomas K. Gaylord, Dr. Andrew F. Peterson, Dr. Jeffrey A. Davis, and Dr. Dennis W. Hess for their time, support, and feedback on my work.

I appreciate the great opportunity to enjoy the vision of my friends and Gigascale Integration group members. I am especially thankful to Dr. Azad Naeemi and Dr. Kaveh Shakeri. Azad was the first one that I was turning to for advice and help on my every question and problem. He and his family, Nassim and Kian, have been like my own family here in Atlanta. I was also lucky that I have a friend like Kaveh with whom I had many discussion and we took many classes together. This work would have not been possible without the help of the other former/graduate students in the GSI group such as Dr. Payman Zarkesh-Ha, Dr. Ragu Venkatesan, Dr. Muhannad Bakir, Deepak Sekar, and Gerald Lopez. I would like to express my gratitude to other GSI members Dr. Azeez Bhavnagarwala, Dr. Keith Bowman, Dr. Tony Mule', Dr. Qiang Chen, Dr. James Joyner, Dr. Raghunath Murali, Dr. Lihui Wang, Dr. Hiren Thacker, Dr. Oluwafemi Ogunsola, Dr. Bing Dang, Gang Huang, Farhana Zaman, and Calvin King.

I would like to acknowledge my other friends in Georgia Tech and Atlanta, in particular, Prof. Ali Adibi, Dr. Fariborz Farahmand, Dr. Pejman Monajemi, Reza Eftekhar, Dr. Arash Karbaschi, Dr. Saman Jafarpour, Ardalan, Dr. Salman Mohagheghi, Dr. Reza Abdolvand, Dr. Nazanin Rahnavard, Dr. Faisal Zaman, Mohammad Soltani, Dr. Babak

Momeni, Dr. Baabak Ashuri, Mohammad Gharipour, Dr. Siavash Pourkamali, Majid Badiei, Saeed Mohammadi, Ehsan Shahhoseini, Farhad Farzbod, Amir Atabaki, Emad Samadiani, Pedram Javanmard, and members of Fusion (2006-2007 champion of Cobb Amateur Soccer League) for making it an enjoyable experience to live here. As I am sure I forgot to mention many names here, my thanks go to everyone who helped me along this way.

I am grateful to all my teachers over the years. Unfortunately, two of my very best teachers in Georgia Tech, Prof. John P. Uyemura and Prof. Kevin F. Brennan, from whom I learned a lot, are not with us anymore. They will be always remembered in the loving memory of their students, in the knowledge that they passed away, and in the great textbooks that they wrote.

GSI group would not be as pleasant as it is without our "GSI mom." I would like to specially thank Jennifer Tatham-Root for taking care of all "GSI kids." Thank you Jennifer!

Last but not least, my love and gratitude go to my parents. If it was not for their love, sacrifices, and supports, I would not be in a position to write this thesis. As teachers, my parents devoted their lives to educate their students, while patiently waited so long to see the last step of their son in his education. Thank you! Thank you! Thank you!

# TABLE OF CONTENTS

Page

# LIST OF TABLES

Page

# LIST OF FIGURES

Page

## SUMMARY

The 2006 International Technology Roadmap for Semiconductors projects that for 2020, interconnects will be as narrow as 14 nm. For a wire with such dimensions, the collision rate of electrons with the surfaces will be a significant fraction of the total number of collisions, causing an increase in the total resistance of the wire. At the same time, interconnections operate at higher frequencies, such that electrons are confined to a sheet near the surface; therefore, the effective cross-sectional area of the conductor decreases and the effective resistance increases. For a wire that operates at ultra-high frequencies, such that skin depth and the mean free path of the electrons are in the same order, skin effect and surface scattering should be considered simultaneously. This is known as the anomalous skin effect (ASE).

The objective of this work is to identify the challenges and opportunities for using GSI interconnects in the nanometer and GHz regime. The increase in the resistivity of a thin wire caused by the ASE is studied. The delay of a digital transmission line resulting from this effect is modeled. Compact models are presented for the bit-rate limit of transmission lines using a general form of resistance that for the first time simultaneously considers dc resistance, skin effect, and surface scattering. A conventional low-loss approximation that is only valid for fast rising signals is also relaxed. In contrast to previous models, it is shown that the bit-rate limit of a transmission line is not scale-invariant. It is also shown that the error of previous models is large (e.g. 80% for bit-rate limit equals reciprocal time-of-flight) if the bit-rate limit is not considerably larger than the reciprocal time-of-flight.

The impact of surface and grain boundary scatterings on the design of multi-level interconnect networks and their latency distribution is studied. For high-performance chips at the 18 nm technology node, it is shown that despite a more than four times increase in the resistivity of copper for minimum-size interconnects, the increase in the number of

metal levels is negligible (less than 7%), and interconnects that will be affected most are so short that their impact on chip performance is inconsequential. It is shown that for low-cost applications where very few wiring pitches are normally used, the number of metal levels needed to compensate for the impact of size effects on the average *rc* delay of a copper interconnect is drastically high. For a case study chip implemented at 14 nm with an interconnect network with two-tiers, it is shown that a 60% increase in the number of metal levels is needed to address a five times increase in the resistivity of minimum-size wires.

As technology advances, it becomes more and more challenging to design on-chip power distribution networks. An optimization methodology has been presented for power distribution interconnects at the local level. For a given IR drop budget, compact models are presented for the optimal widths of power and ground lines in the first two metal levels for which the total metal area used for power distribution is minimized. Wire widths and thicknesses at the end of the 2006 ITRS are projected to scale down to 14 nm, and size effects are expected to increase copper resistivity by more than four times. Either a three times increase in wiring area for local power lines or a two times decrease in the power via pitch is necessary to compensate for size effects. However, decreasing the power via pitch will increase the power via blockage factor for all metal levels between Metal 2 and the top-most levels with global power grids. The results shows that for 2020, the optimal design that minimizes both the area needed for the local power distribution and power via blockage factor, occurs when power via pitches in the $x$ and $y$ directions are equal and widths of the power lines in Metal 1 are three times the minimum size wires in Metal 1.

# CHAPTER 1

# Introduction

Since the invention of the IC (integrated circuit) by Jack Kilby in 1958, nothing has remained the same except for the incredible rate at which the size of devices is shrinking. This trend, known as Moore's law, has been translated into an exponential increase in the performance of very large scale integration (VLSI) chips in the past three decades [1]. Arguably, this law is expected to hold for another decade. The two main physical challenges facing Moore's law are "power dissipation/heat removal" and "interconnects." Like transistors, the cross-section of metal wires is scaled down with every generation, but, unlike transistors, wire performance drops as they become thinner, due to increased resistance [2, 3]. In deep submicron designs, the interconnect delay exceeds the device delay and is the dominant factor of chip performance. Typically, transmitting data between different chip components requires a few clock cycles. Reverse scaling in the multi-level interconnect network is used to ease this problem [4].

About a decade ago, copper replaced aluminum as an advanced metallization solution [5]. As the feature sizes of interconnects scale down, copper resistivity increases as a result of various size effects, such as *surface scattering*, *grain boundary scattering*, *surface roughness*, and *barrier thickness*. The mean free path of electrons in bulk copper at room temperature is about 39 nm [6]. The 2006 ITRS projects that at the end of the roadmap, interconnects will be as narrow as 14 nm [7]. For a wire with such dimensions, the collision rate of electrons with the sidewall surfaces and grain boundaries will be a significant fraction of the total number of collisions, causing an increase in the total resistance of the wire.

At the same time, interconnections operate at higher frequencies, such that electrons are confined to a sheet near the surface known as the "skin depth." The projected clock

frequency for 2018 will be 53 GHz, and the skin depth at the third harmonic of the clock frequency will be as small as 160 nm. Hence, the skin depth becomes comparable to the mean free path of the electrons; therefore, skin effect and surface scattering should be considered simultaneously. This is known as the anomalous skin effect (ASE). Figure 1.1 shows wire dimension and skin depth at the $3\omega$ frequency versus different technology nodes based on ITRS projections and are compared with the mean free path of electrons in copper ($\lambda_{Cu}$) at room temperature.

*The main objective of this thesis is to understand the problems and challenges that are caused by the trend shown in Figure 1.1 and to study possible solutions and opportunities for copper wires as gigascale integration (GSI) interconnects.*

The outline of this thesis is as follows. In Chapter 2, a new model for the ASE is derived. In this model, first a closed-form solution for the current density and conductivity of thin wires assuming an exponential dependency for the electric field is developed. Then, the closed-form solution combined with the Fourier exponential series of an arbitrary electric field is used to find the accurate physical model of conductivity and current distribution of thin wires under the ASE. In Chapter 3, a new methodology for modeling a transmission line in its most general distributed circuit form is proposed. This method is used to find the bit-rate limit formula for copper wires. The impact of the ASE on delay and bit-rate limit is studied. Chapter 4 describes the impact of size effects on design and performance of a multi-level interconnect network of both a high-performance and low-cost application. The size effect and its consequent impact on the optimal design of a local power distribution network is presented in Chapter 5. Finally, Chapter 6 provides the conclusions and directions for future research.

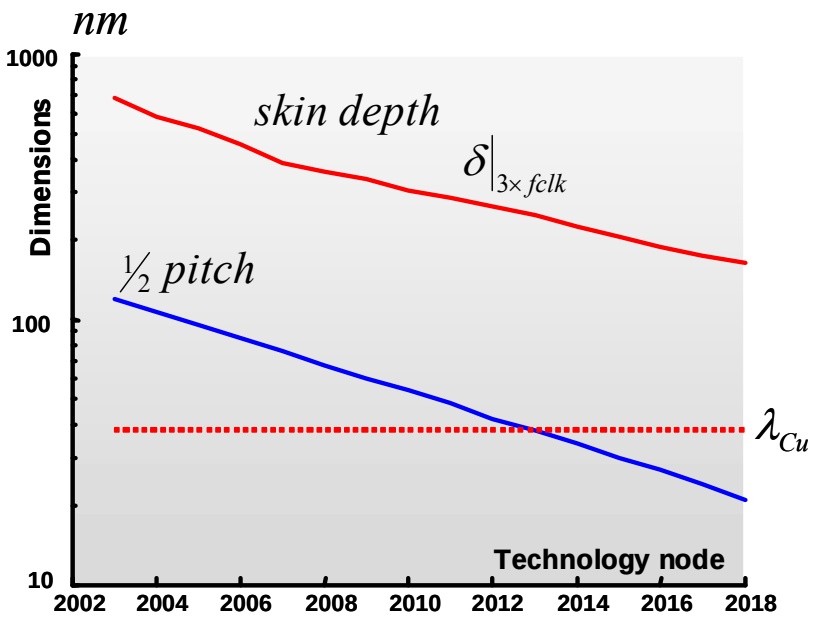



Figure 1.1: Wire dimension (½ pitch), skin depth at the $3\omega$ frequency ($\delta|_{3\times fclk}$), and mean free path of electrons in copper at room temperature ($\lambda_{Cu}$) versus different technology nodes based on ITRS projections.

## 1.1 Size Effects and ASE Modeling

Copper will be the preferred interconnect material for GSI applications for at least the next decade. The continuous shrinking in feature size will impose new issues for copper wires, including size effects. The size-affected wires may have prohibitively high resistivity, thus affecting chip functionality. When $\lambda$ (mean free path of electrons, defined as the mean distance between two consecutive phonon scatterings) becomes comparable to the wire dimensions, a significant portion of conducting electrons loses their momentum during interaction with the surface. This is known as surface scattering.

Low-resistivity inlaid copper interconnects for GSI applications are formed by using a copper damascene process [8]. In this process, the surface of interlayer dielectric film is patterned by standard photolithography and reactive ion etching. Then, copper is deposited into the recessed trench by electroplating, and excess copper is removed by

chemical mechanical polishing. The electroplated copper is formed with different grain boundaries. An electron traveling across the grain boundaries will be scattered from the boundaries, this is known as grain boundary scattering. The thickness of barrier layers and the roughness of the copper surface will also lead to an increase in the effective resistivity of the wires [9].

The anomalous skin effect was observed by Pippard while he was studying the Fermi surface of copper under the action of an electromagnetic field [10]. The electromagnetic field interacts with the electrons of the metal, hence electrons absorb energy from the field. The nature of interaction depends on the ratio of $\delta$ (skin depth, defined as the distance in which the amplitude of the wave falls to 1/*e* of its value at the surface) to $\lambda$. When $\lambda << \delta$, the normal skin effect region, the impact of the mean free path of the electrons can be neglected. Therefore, the relation between the current vector and the electric field vector might be considered by a scalar value known as metal conductivity. However, when $\lambda > \approx \delta$, the effect on the electrons that move perpendicular to the surface is much different from those traveling parallel to the surface. Due to the long mean free path, electrons in the first case leave the skin depth region layer without being scattered. Thus, only those electrons that are moving approximately parallel to the surface contribute to the absorption. One of the applications of the ASE is as one of the techniques for measuring Fermi surface [11].

Previous models for size effects and ASE have been derived [10, 12-14]. Based on these models in terms of $w/\lambda$ (ratio of wire dimension to electron mean free path) and $\delta/\lambda$ (ratio of classical skin depth to electron mean free path) four different regions can be defined: (1) bulk dc resistance, (2) size-effect limited dc resistance, (3) classical skin effects, and (4) ASE. Asymptotic formulas for copper resistivity in each of these regions are known. Figure 1.2 indicates these regions approximately. Historically, size effects and ASE have been studied at low-temperature physics. Since the mean free path of electrons is larger at lower temperatures, size effects and ASE occur at larger dimension

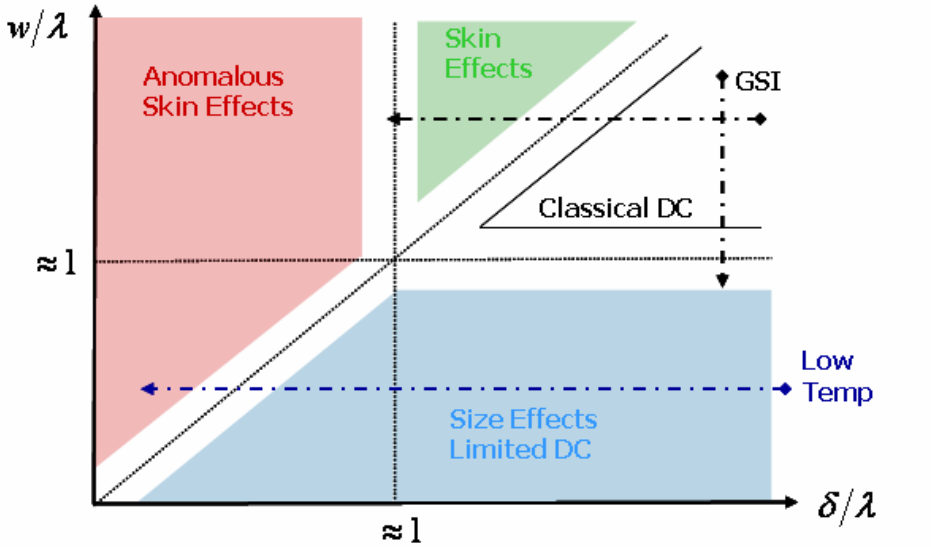



Figure 1.2: Indication of regions of validity for various asymptotic formulas for the impedance of copper wires in terms of $w/\lambda$ (ratio of wire dimension to electron mean free path) and $\delta/\lambda$ (ratio of classical skin depth to electron mean free path).

and smaller frequencies. However, for GSI application, by shrinking to dimensions around 14 nm, we leave the dc bulk resistivity region and enter the size-effect-limited region, and by increasing the clock frequency up to 50 GHz, we leave the classical skin effect region, although we do not enter the ASE region until reaching frequencies above 300 GHz for copper wires at room temperature. Hence, it is important to have a model for the frequency dependency of copper wires at room temperature that covers the classical skin effect, ASE region, and the region in between.

## 1.2 Phasor Analysis

Maxwell's equations for TEM waves on transmission lines reduce to the Telegrapher's equations. Mostly, the four parameters for the distributed model of TEM transmission

lines are series resistance, inductance per unit length, shunt conductance, and capacitance per unit length. It is shown that Telegrapher's equations are valid for higher frequencies, while inherently longitudinal field components exist. This is known as the quasi-TEM approximation [15], while the transmission line parameters are frequency dependent. The most general way to model a transmission line as a distributed network is to consider a general series impedance per unit length as $z(s)$ and a general parallel admittance per unit length as $y(s)$. The objective of this section is to find a method called "phasor analysis" to solve a distributed transmission line in its most general form. Such a methodology is needed to study the step response and bit-rate limit of electrical interconnects considering size effects and high frequency effects.

### 1.2.1 Bit-Rate Limit Model

As digital circuits approach gigascale integration, the demand for transmitting data at very high rates grows rapidly. Electrical interconnects have been conventionally used for on-chip and chip-to-chip interconnection, and because of the simplicity of the transmitters and receivers they need, they will remain an attractive solution for GSI interconnection. As mentioned in the previous section, various effects, including surface scattering and skin effect, affect electrical interconnects and make it difficult to accurately determine the step response and the bit-rate limit of electrical wires. Physical models that accurately determine the bit-rate limit of electrical wires are very helpful in determining their capabilities at various generations. Such models can also help optimize the cross-sectional dimensions of wires to maximize aggregate bandwidth.

Previous models for wire bit-rate have several major limitations [16-18]. They are not general because they have been derived for specific structures, like coaxial cables or striplines, while this new proposed model is based on transmission line parameters instead of geometrical parameters like the wire's cross-sectional area or perimeter. In previous models, the dc resistance of wires has also been neglected, causing an overesti-

mation of the bit-rate limit if the signal rise time is not small enough. Generally, all of the previous models use low-loss approximation for a transmission line, which can cause errors when the bit-rate limit is not considerably larger than the reciprocal time-of-flight ($T_F$) [19].

## 1.3 Impact of Size Effects on Multi-Level Interconnect Network

As the number of transistors per GSI chip increases, the total length of wires on chip also increases. For high-performance chips, these wires are routed in several different tiers. Based on their length (and pitch sizes), the wires can be separated into local, intermediate, and global wires. Local wires connect gates and transistors within a functional block and are usually routed in the minimum pitch and occupy the first two metal levels of a multi-level interconnect network. The lengths of these wires are usually less than a few gate pitches so their length scales down with the technology. Intermediate wires provide clock and signal distribution within a functional block or inter-module communications between adjacent blocks with typical lengths up to 3~4 mm. Global wires provide clock and signal distribution between functional blocks and deliver power/ground to all functions on a chip. Global wires that are routed in the top metal levels are longer than 4mm and can be as long as the chip size. As the length of global wires does not scale down with the technology, the overall performance of the interconnect network can become dominated by global wires. The insertion of repeaters can mitigate this problem but is not enough. In addition, reverse scaling of global and intermediate wires is necessary.

Typical interconnect length distribution and a chip cross-section using reverse scaling for its multi-level interconnect network are shown in Figure 1.3. Knowing the wiring distribution to estimate the interconnect lengths a priori [20, 21], the reverse-scaled multi-level interconnect networks can be optimized to reduce the logic macro-cell area, cycle time, power consumption, or number of metal levels [4]. These considerations are

unavoidable for GSI chips where interconnects are the limiting factor. Optimization of a multi-level interconnect network is a part of the interconnect-centric design paradigm [22].

Size effects exacerbate the interconnect problems. The resistivity of minimum-size wires at the end of the ITRS roadmap could be five times more than the bulk resistivity. However, a multi-level interconnect network consists of interconnects with a wide range of lengths that are routed in metal levels with different pitches and thicknesses. The impact of size effects on the design of multilevel interconnect networks has not been studied, and it is not clear how much overall chip performance will be degraded because of size effects. Such quantitative studies can be very helpful for any kind of cost/performance analysis for various size-effect mitigation techniques.

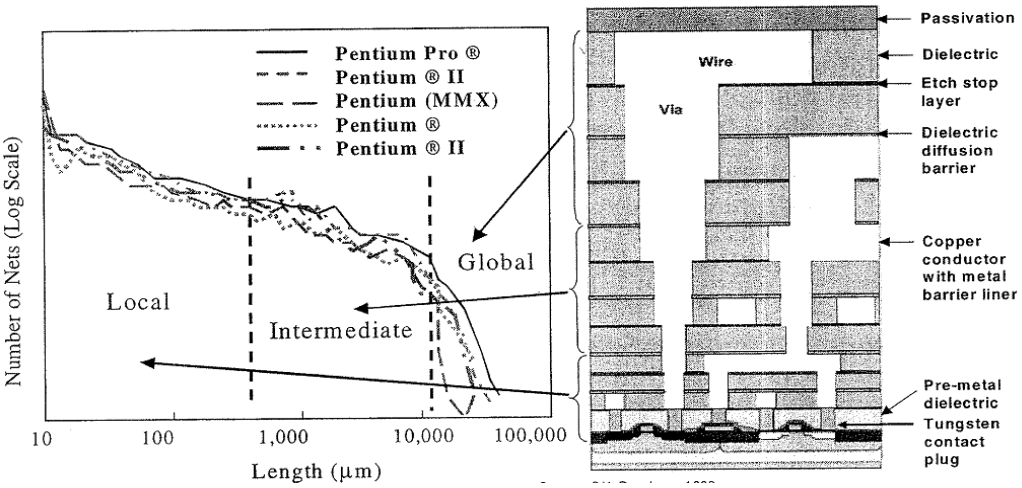



Figure 1.3: Typical interconnect length distribution (left) and chip cross section (right) with reverse interconnect scaling to alleviate interconnect problems [22].

## 1.4 Impact of Size Effects on Local Power Distribution Network

As technology advances, it becomes more and more challenging to design on-chip power distribution networks. Previously, making a wire wide enough to limit the IR voltage drop was enough for designing the power distribution network. Currently, with the increased density and speed of GSI chips, power distribution becomes complex. Different topologies such as power ring, rooted and non-rooted trees, solid planes, and meshed grids are employed for power networks [23-25]. Different techniques have been used to minimize the loop inductance and $Ldi/dt$. To ensure reliable operation of the power distribution system, accurate simulation and synthesis of the power distribution network over the whole frequency range are essential, although the resistivity per unit length of the local power distribution network is so high that inductive effects are negligible. Therefore, for the local power distribution network the $IR$ drop and for the global power distribution network both the $IR$ drop and $Ldi/dt$ should be considered in the modeling and optimization of a power network.

At the global level, power and ground lines are routed in the top metal levels with relatively large cross-sectional dimensions, and size effects do not affect them. At the local level, however, power and ground lines are routed in the first two metal levels to deliver power and ground to each individual gate. Due to their small dimensions, these interconnects are vulnerable to size effects. It is therefore critical to model and optimize power distribution networks at the local level and to quantify the impact of size effects on the design and performance of local power distribution networks, which are the subject of this section.

# CHAPTER 2

# ELECTRONS IN METALS

## 2.1 Introduction

About a decade ago, copper replaced aluminum as an advanced metallization solution and it will be the preferred interconnect material for GSI applications for at least the next decade. This transition was driven in part by copper's relatively low resistivity. However, as the wire dimensions approach the mean free path of electrons for copper (39 nm at room temperature), copper resistivity increases as a result of various size effects such as surface scattering, grain boundary scattering, surface roughness, and barrier thickness [9].

In the simplest model, metal resistivity is caused by collisions of the conduction electrons with phonons, and mean free path is defined as the average distance that an electron travels between two subsequent collisions. The total resistivity of metal is a combination of the contribution of two individual and independent scattering mechanisms, thermal resistivity and defect resistivity (Matthiessen's rule). Electron collisions with vibrating atoms displaced from their equilibrium lattice positions are the source of the thermal or phonon contribution, which increases linearly with temperature. Impurity atoms, defects such as vacancies, and grain boundaries locally disrupt the periodic electric potential of the lattice and effectively cause electron scattering, which is temperature independent. Figure 2.1 illustrates the surface scattering and grain boundary scattering of a copper interconnect [26].

When the critical dimension of the copper wires is smaller than the electron mean free path, the electrons' motion will be interrupted by collisions with the surface. The electrons undergo either elastic or inelastic scattering. In elastic, also known as specular,

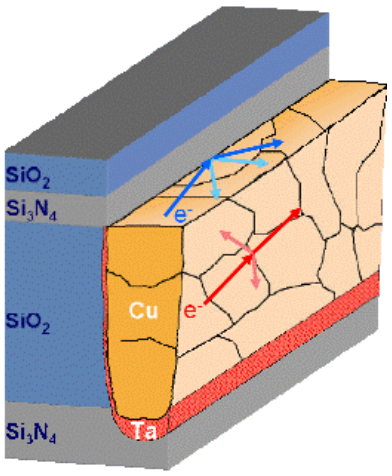



Figure 2.1: Schematic illustration of surface and grain boundary scattering for copper interconnects with Tantalum (Ta) barrier [9].

scattering, the electron reflects in the same way a photon reflects from a mirror. In this case, the electron does not lose its energy and its momentum or velocity along the direction parallel to the surface is preserved. As a result, the electrical conductivity remains the same as in the bulk and there is no size effect on the conductivity. When scattering is totally inelastic or nonspecular, the electron mean free path is terminated by impinging on the surface. The scattered electron loses its velocity along the direction parallel to the surface or the conduction direction, and the electrical conductivity decreases. There will be a size effect on electrical conduction.

At the same time, interconnections operate at higher frequencies, such that electrons are confined to a sheet near the surface known as "skin depth." The projected clock frequency for the year 2018 is 53 GHz, and the skin depth at the $3\omega$ frequency will be as small as 160 nm. Hence, the skin depth becomes comparable to the mean free path of the

electrons; therefore, skin effect and surface scattering should be considered simultaneously.

The physics of conduction in thin wires at high frequencies is different from that of bulk conductivity. Previous studies on ASE have been done for wires at high frequencies and very low temperatures [14, 27, 28]. A physical model for the ASE of copper wires at room temperature and ultra-high frequencies is needed to study the behavior of metallic wires for GSI interconnects. Two different approaches have been used to model the conductivity of thin wires. The first approach, which is more qualitative, is based on the energy relaxation of electrons and has been used by Chambers [12]. The second approach, which is more detailed, uses the Boltzmann transport equation to find the distribution of electrons in the *k*-space and has been used by Sondheimer [14]. In this work, we will use the Sondheimer approach to model size effects and will extend it to a model for the ASE.

In this chapter, after a brief review of the theory of electrons in metals, a closed-form solution for the current density and conductivity of thin wires is developed, assuming an exponential dependency for the electric field. This closed-form solution combined with the Fourier exponential series of an arbitrary electric field will enable us to find the accurate physical model of conductivity and current distribution of thin wires under the ASE [29]. Such models can be used to study the potential and opportunities of copper interconnects compared with other alternative potential solutions for GSI interconnects [30]. Finally, a simple model for resistivity of metallic wires with line-edge roughness is proposed.

## 2.2 Theory of Electrons in Metals

### 2.2.1 Background

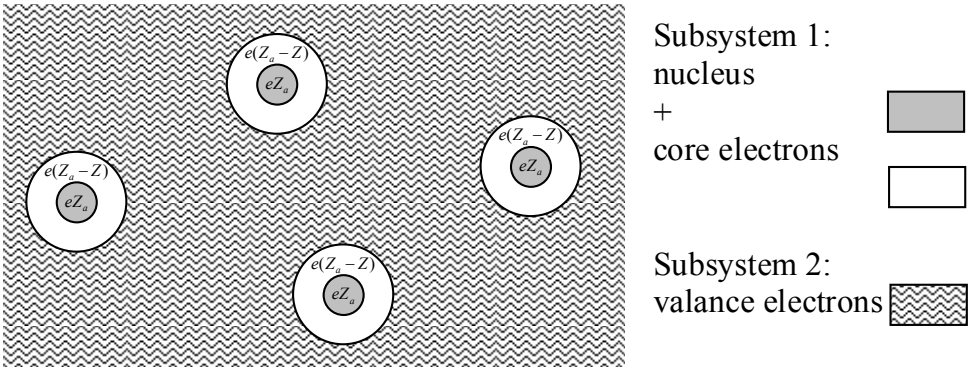



Figure 2.2: Schematic indication of D-L-S model. A metal is divided into two different subsystems: core electrons, valence electrons

.

At the beginning of the 20th century, along with many other advances in science, the "modern electron theory of metals" was developed. In 1900, three years after Thompson's discovery of the electron, Drude constructed his theory of electrical and heat conduction by applying the highly successful kinetic theory of gases to a metal, considered a gas of free electrons. Afterward, Lorentz resolved the behavior of electrons by means of the statistical methods of the dynamical theory of gases. The Drude-Lorentz model fails to explain why the conduction electrons do not contribute appreciably to the "specific heat of a metal." This problem was solved later when Sommerfeld (and Pauli) applied Fermi-Dirac statistics to the free electrons in the model presented by Drude-Lorentz for a metal [31].

In the Drude-Lorentz-Sommerfeld (D-L-S) model, a metal is divided into two different subsystems, as shown in Figure 2.2. The core electrons are attached to the heavy nucleus, forming immobile metal ions, while valence electrons with mass *m* are allowed to wander around.

The assumptions made in this model are as follows:

(1) All the valence electrons are free to move throughout the metal.

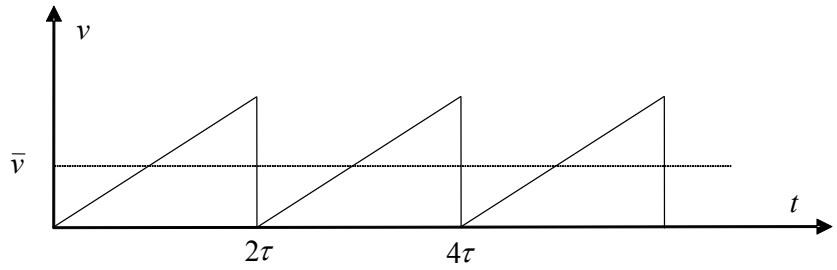



Figure 2.3: Electron behavior in D-L-S model

(2) Electrons started at rest and were accelerated by an applied electric field until they suffered a collision in time $2\tau$.

(3) Electrons collide with ion cores.

Based on these assumptions, as shown in Figure 2.3, the average drift velocity of electrons will be

$$\bar{v} = \frac{v_m}{2} = \frac{eE\tau}{m}, \tag{2.1}$$

where $e$ is electron charge, $E$ is applied electric field, $m$ in the electron mass, and $\tau$ is electron relaxation time. The current density of electrons is given by

$$J = ne\bar{v} = \frac{ne^2\tau}{m}E = \sigma E, \tag{2.2}$$

where $\sigma$ is called the bulk conductivity, and $n$ is the number of free electrons per unit volume, and is on the same order as the number of atoms per unit volume, given by

$$n = \frac{8\pi}{3}\left(\frac{mv_F}{h}\right)^3, \tag{2.3}$$

where $v_F$ is the electron velocity at the Fermi surface [32]. The relaxation time is related to the electron velocity on the Fermi surface and the bulk mean free path (m.f.p. $\lambda$) by

$$\lambda = v_F\tau\,. \tag{2.4}$$

In a more accurate term, $\tau$ is defined in such a way that, the probability $P$ of an electron existing for a time $t$ without being scattered is written by the rate equation [33]

$$\frac{\partial P}{\partial t} = -\frac{P}{\tau}. \tag{2.5}$$

$\tau$ depends on the electron energy $E$ as

$$\tau = \tau_b E^q\,, \tag{2.6}$$

where $\tau_b$ is independent of $E$, and $q$ takes values that depend on the scattering mechanism; $q = -1/2$ in the case of lattice scattering and $q = 3/2$ in the case of ionized impurity scattering [34]. In the framework of quantum mechanics, the simple D-L theory remains valid when the electron is regarded as a free particle with an effective mass $m^*$.

### 2.2.2 Boltzmann Equation for Charge Transport

One way to describe the behavior of electrons is by means of a distribution function $f(\vec{v},\vec{r})$, which measures the number $dN(\vec{v})$ of electrons with velocity $\vec{v}$ lying in the range $d\vec{v}$ in the volume element $d\vec{r}$, i.e.,

$$dN(\vec{v}) = 2\left(\frac{m^*}{h}\right)^3 f(\vec{v},\vec{r})d\vec{r}d\vec{v}\,, \tag{2.7}$$

in the absence of any perturbation [32],

$$f(\vec{v},\vec{r}) = f_0(v) \tag{2.8}$$

where $f_0(v)$ is the Fermi-Dirac distribution, i.e.,

$$f_0(v) = 1\Big/(1+\exp(\frac{m}{2kT}(v^2 - v_F^2))) . \tag{2.9}$$

where $k$ is Boltzmann's constant, and $T$ is temperature.

Three different mechanisms contribute to the non-equilibrium distribution function: drift, diffusion, and collisions [27].

*(I) Drift*: $\vec{E}$ and $\vec{B}$, electric and magnetic fields, induce change in the velocity vector of free electrons as (dot over the variable denote the time derivative $\dot{x} = dx/dt$)

$$\dot{\vec{v}} = -\frac{e}{m^*}(\vec{E} + \vec{v} \times \vec{B}) . \tag{2.10}$$

and consequently a change in the distribution function as

$$\left.\frac{\partial f}{\partial t}\right|_{drift} = -\dot{\vec{k}} \cdot \vec{\nabla}_k f = \frac{e}{\hbar}(\vec{E} + \vec{v} \times \vec{B}) \cdot \vec{\nabla}_k f . \tag{2.11}$$

where $f$ stands for $f(\vec{v},\vec{r})$ for simplicity and $\vec{\nabla}_k f$ denotes the k-space gradient of the distribution function (in Cartesian coordinates $\vec{\nabla}_k = \frac{\partial}{\partial k_x}\hat{k}_x + \frac{\partial}{\partial k_y}\hat{k}_y + \frac{\partial}{\partial k_z}\hat{k}_z$).

*(II) Diffusion:* A gradient in the spatial distribution of electrons will cause a change in the distribution density at the rate of

$$\left.\frac{\partial f}{\partial t}\right|_{diffusion} = -\dot{\vec{r}} \cdot \vec{\nabla}_r f . \tag{2.12}$$

where $\vec{\nabla}_r f$ denotes the gradient of the distribution function (in Cartesian coordinates $\vec{\nabla}_r = \frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial y}\hat{j} + \frac{\partial}{\partial z}\hat{k}$).

*(III) Collisions:* Suppose a non-equilibrium distribution function is set up by a system of external forces that are suddenly removed; the rate of approach to equilibrium under the influence of collisions alone (with phonons and imperfections) is then given by

$$\left.\frac{\partial f}{\partial t}\right|_{collision} = -\frac{f - f_0}{\tau} . \tag{2.13}$$

Combining these three scattering mechanisms, the total contribution to the distribution function can be written as

$$\begin{aligned} \frac{\partial f}{\partial t} &= \left.\frac{\partial f}{\partial t}\right|_{drift} + \left.\frac{\partial f}{\partial t}\right|_{diffusion} + \left.\frac{\partial f}{\partial t}\right|_{collision} = \\ &= \frac{e}{\hbar}(\vec{E} + \vec{v} \times \vec{B}) \cdot \vec{\nabla}_k f - \vec{\dot{r}} \cdot \vec{\nabla}_r f - \frac{f - f_0}{\tau} \end{aligned} . \tag{2.14}$$

where in steady-state conditions (i.e. $\partial f / \partial t = 0$), the Boltzmann equation for quasi-free electrons can be written as

$$-\frac{e}{\hbar}(\vec{E} + \vec{v} \times \vec{B}) \cdot \vec{\nabla}_k f + \vec{\dot{r}} \cdot \vec{\nabla}_r f = -\frac{f - f_0}{\tau} . \tag{2.15}$$

which is purely classical, except that the mass *m* is to be regarded as an effective mass. The main problem for solving problems such as surface scattering, grain boundary scattering, surface roughness, and anomalous skin effects, then, is to solve equation (2.15) for various cases of interest with its boundary conditions. The boundary conditions are determined by the nature of the scattering and physical boundaries of the specimen. The solution of *f* will be used to calculate the current density (in $A/m^3$) by means of the usual formula [32]:

$$\vec{J} = -2e\left(\frac{m^*}{h}\right)^3 \int \vec{v} f \, d\vec{v} . \tag{2.16}$$

## 2.3 Conductivity of Thin Film

In this section strict analysis of the conductivity of a thin metallic film is given in detail. As this method is typical of all calculations in different cases (such as thin wires, anomalous skin effect), it is presented in detail. This method first was used by Sondheimer [14].

Consider a metal film of thickness $d$ and suppose that the $z$-axis is perpendicular to the plane of the film, with film surfaces at $z$=0 and $z$=$d$ (Figure 2.4). Now, the problem is one dimensional, and the distribution function of the electrons may be written in the form

$$f = f_0 + f_1(\vec{v}, z)\,, \tag{2.17}$$

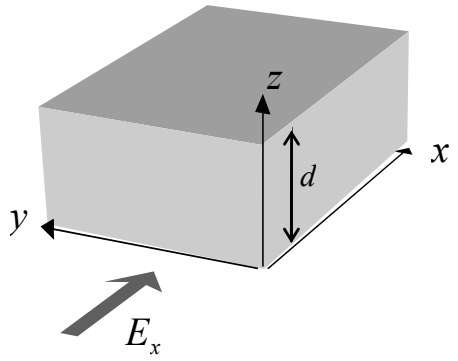



Figure 2.4: Definition of coordinates for a metal film

where $f_1$ has to be determined. The electric field $E$ is in the $x$-direction and the Boltzmann equation (2.15) reduces to

$$\frac{f_1}{\tau} + v_z \frac{\partial f_1}{\partial z} = \frac{eE_x}{m^*} \frac{\partial f_0}{\partial v_x}\,, \tag{2.18}$$

The general solution is easily written and will be

$$f_1(\vec{v}, z) = \frac{e\tau E_x}{m^*} \frac{\partial f_0}{\partial v_x} \left\{ 1 + F(\vec{v}) e^{-z/\tau v_z} \right\}, \tag{2.19}$$

where $F(v)$ is an arbitrary function of $v$. To determine $F(v)$, we have to introduce the boundary conditions at the surface of the film. We assume that a fraction $p$ of the electrons is scattered elastically at the surface with reversal of the velocity component $v_z$, while the rest are scattered diffusely with complete loss of their drift velocity. The value

of $p$ varies between 0 and 1, where $p$=1 means specular scattering and $p$=0 means complete diffuse scattering (Figure 2.5). Specular reflection of waves requires that surface irregularities be smaller than the wavelength. In metals, the de Broglie wavelength of electrons on the Fermi surface is of the order of one inter-atomic distance. Hence, even the most carefully prepared macroscopically "flat" surface fails to meet such a stringent requirement. It is interesting to note that specular reflection has been observed in bismuth films. The number of conduction electrons in bismuth, a semimetal, is extremely small; thus the de Broglie wavelength of electrons is several hundred inter-atomic distances [35].

Assuming two distribution functions, $f^+$ for electrons with $v_z > 0$ and $f^-$ for electrons with $v_z < 0$, the distribution function of the electrons leaving the surface $z$=0 is now given by

$$f_0 + f_1^+(v_z, z=0) = p\left\{f_0 + f_1^-(-v_z, z=0)\right\} + (1-p)f_0, \qquad (2.20,a)$$

and similarly, at $z$=$d$,

$$f_0 + f_1^-(v_z, z=d) = p\left\{f_0 + f_1^+(-v_z, z=d)\right\} + (1-p)f_0, \qquad (2.20,b)$$

The solution of (2.19) with these boundary conditions will be

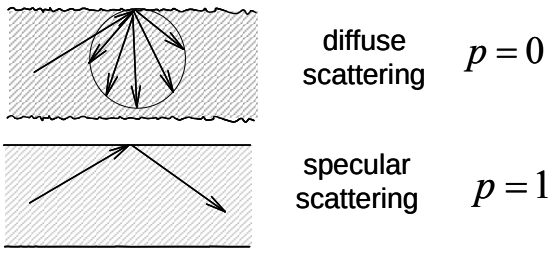



Figure 2.5: Specularity parameter is defined as the fraction of electrons that have elastic collisions at the wire surfaces

$$\left.\begin{aligned} f_1^{+}(\vec{v},z) &= \frac{e\tau E_x}{m^*}\frac{\partial f_0}{\partial v_x}\left\{1-\frac{1-p}{1-pe^{-d/\tau v_z}}e^{-z/\tau v_z}\right\} \qquad (v_z>0) \\ f_1^{-}(\vec{v},z) &= \frac{e\tau E_x}{m^*}\frac{\partial f_0}{\partial v_x}\left\{1-\frac{1-p}{1-pe^{d/\tau v_z}}e^{(d-z)/\tau v_z}\right\} \qquad (v_z<0) \end{aligned}\right\}. \tag{2.21}$$

Combining (2.16), (2.17), and (2.21) we can now proceed with the calculation of current density $J(z)$.

$$J(z) = -2e(\frac{m^*}{h})^3 \frac{e\tau E_x}{m^*}\left\{\begin{aligned} &\iiint_{v_z>0} v_x \frac{\partial f_0}{\partial v_x}\left\{1-\frac{1-p}{1-pe^{-d/\tau v_z}}e^{-z/\tau v_z}\right\}dv_x dv_y dv_z + \\ &\iiint_{v_z<0} v_x \frac{\partial f_0}{\partial v_x}\left\{1-\frac{1-p}{1-pe^{d/\tau v_z}}e^{(d-z)/\tau v_z}\right\}dv_x dv_y dv_z \end{aligned}\right\} \tag{2.22}$$

Introducing spherical coordinates ($v$, $\theta$, $\varphi$), as shown in Figure 2.6, the Cartesian integral in (2.22) can be written in the spherical coordinate using the relation

$$\iiint f(x,y,z)dxdydz = \iiint r^2 \sin\theta\, f(r,\theta,\varphi)\, dr d\theta d\varphi\,, \tag{2.23}$$

as

$$\begin{aligned} J(z) = -2e(\frac{m^*}{h})^3 \frac{e\tau E_x}{m^*}\int_0^\infty dv \int_0^{2\pi} d\varphi\, v^3 \cos^2\varphi \frac{\partial f_0}{\partial v_x} \\ \times\left[\begin{aligned} &\int_0^{\pi/2} \sin^3\theta\left\{1-\frac{1-p}{1-pe^{-d/\tau v\cos\theta}}e^{-z/\tau v\cos\theta}\right\}d\theta \\ &+\int_{\pi/2}^{\pi} \sin^3\theta\left\{1-\frac{1-p}{1-pe^{d/\tau v\cos\theta}}e^{(d-z)/\tau v\cos\theta}\right\}d\theta \end{aligned}\right]. \end{aligned} \tag{2.24}$$

We used the fact that $f_0$ is as shown in (2.9). Therefore,

$$\frac{\partial f_0}{\partial v_x} = \frac{v_x}{v}\cdot\frac{\partial f_0}{\partial v}, \tag{2.25}$$

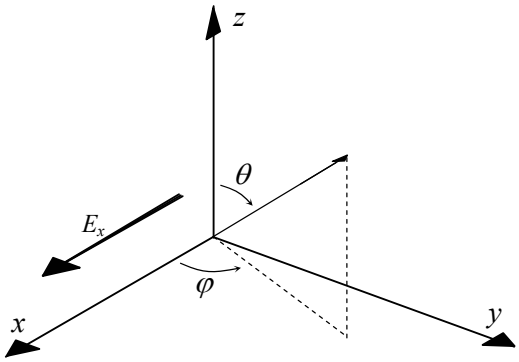



Figure 2.6: Parameter definitions in spherical and Cartesian coordinates

In (2.24) the integration over $\varphi$ is trivial, and the integration over $v$ is carried out by means of degenerate electron gas of (2.9) as

$$-\int_0^\infty \psi(v)\frac{\partial f_0}{\partial v}dv = \psi(v_F)\,. \tag{2.26}$$

Therefore,

$$J(z) = -\frac{2\pi e^2 m^{*2}\tau v_F^3}{h^3}E\int_0^{\pi/2}\sin^3\theta\left\{2-\frac{(1-p)e^{-z/\lambda\cos\theta}}{1-pe^{-d/\lambda\cos\theta}}-\frac{(1-p)e^{(z-d)/\lambda\cos\theta}}{1-pe^{-d/\lambda\cos\theta}}\right\}d\theta\,, \tag{2.27}$$

where $\lambda=\tau v_F$ is the free path of the electrons at the surface of the Fermi distribution. The overall conductivity of the film could be found by averaging the current density over all values of $z$ from 0 to $d$ to obtain

$$\begin{aligned}\sigma &= \frac{1}{Ed}\int_0^d J(z)dz \\ &= \frac{4\pi e^2 m^{*2}\tau v_F^3}{h^3}\left[\int_0^{\pi/2}\sin^3\theta\,d\theta-\frac{(1-p)\lambda}{d}\int_0^{\pi/2}\sin^3\theta\cos\theta\frac{1-e^{-d/\lambda\cos\theta}}{1-pe^{-d/\lambda\cos\theta}}d\theta\right], \\ &= \sigma_0\left(1-(1-p)\frac{3\lambda}{2d}\int_0^{\pi/2}\sin^3\theta\cos\theta\frac{1-e^{-d/\lambda\cos\theta}}{1-pe^{-d/\lambda\cos\theta}}d\theta\right)\end{aligned} \tag{2.28}$$

where

$$\sigma_0 = \frac{8\pi}{3}\frac{e^2 m^2 \tau v_F^3}{h^3}, \tag{2.29}$$

## 2.4 Conductivity of Thin Wire

The Boltzmann transport equation (2.15) for a thin wire will be simplified as

$$\frac{f_1}{\tau} + v_z \frac{\partial f_1}{\partial z} + v_y \frac{\partial f_1}{\partial y} = \frac{eE_x}{m^*}\frac{\partial f_0}{\partial v_x}. \tag{2.30}$$

Now, $f_1$ is both a function of $z$ and $y$ and the integral in (2.22) should be done in four different regions. Suppose that the line has a width of $w$ in the $z$ direction and a height of $h$ in the $y$ direction. The boundary condition for the BTE (2.30) will be (for simplicity assume $p$=0):

$$\begin{aligned}
&\text{Region1:}\begin{cases}\theta\in[0,\pi/2],\varphi\in[0,\pi]\\ v_z>0, v_y>0\end{cases}\rightarrow\begin{cases}f(v,0,y)=0\\ f(v,z,0)=0\end{cases}\\
&\text{Region2:}\begin{cases}\theta\in[0,\pi/2],\varphi\in[\pi,2\pi]\\ v_z>0, v_y<0\end{cases}\rightarrow\begin{cases}f(v,0,y)=0\\ f(v,z,h)=0\end{cases}\\
&\text{Region3:}\begin{cases}\theta\in[\pi/2,\pi],\varphi\in[0,\pi]\\ v_z<0, v_y>0\end{cases}\rightarrow\begin{cases}f(v,w,y)=0\\ f(v,z,0)=0\end{cases}\\
&\text{Region4:}\begin{cases}\theta\in[\pi/2,\pi],\varphi\in[\pi,2\pi]\\ v_z<0, v_y<0\end{cases}\rightarrow\begin{cases}f(v,w,y)=0\\ f(v,z,h)=0\end{cases}
\end{aligned}. \tag{2.31}$$

Using the same technique as in the previous section, the resistivity will be

$$\begin{aligned}
\frac{\rho_0}{\rho} = 1 &- \frac{6}{4\pi hw}\int_0^w dz\int_0^h dy\int_{-\tan^{-1}(y/z)}^{\tan^{-1}(h-y/z)} d\varphi\int_0^\pi d\theta\left\{\sin\theta\cos^2\varphi\exp\left(\frac{-z}{\lambda\sin\theta\cos\varphi}\right)\right\}\\
&- \frac{6}{4\pi hw}\int_0^w dz\int_0^h dy\int_{-\tan^{-1}(z/h-y)}^{\tan^{-1}(w-z/h-y)} d\varphi\int_0^\pi d\theta\left\{\sin\theta\cos^2\varphi\exp\left(\frac{y-h}{\lambda\sin\theta\cos\varphi}\right)\right\}
\end{aligned}. \tag{2.31}$$

Throughout this work we define resistivity by

$$\rho = \frac{A}{l}R. \tag{2.32}$$

where $A[m^2]$, $l[m]$, and $R[\Omega]$ are the cross-sectional area, length, and the resistance of the sample material, respectively. Based on this definition, the resistivity of a thin wire with height $h$ and width $w$ is plotted in Figure 2.7. These curves are plotted assuming $p$=0 and for copper wires with different aspect ratios, and compared with thin film (case $h/w=\infty$). Although the general theory seems complicated, some simple interpretations have been done before. For a wire with arbitrary cross-section and thickness larger than or comparable to the mean free path, the deviation in the current density from that which would exist in the bulk is appreciable only in a region close to the surface of the wire. Therefore, the deviation in the current is proportional to the perimeter of the wire's cross-section $P$, while mean current is proportional to the cross-sectional area $A$. Simply, for a wire with arbitrary cross-section the thin wire conductivity can be written as [13]

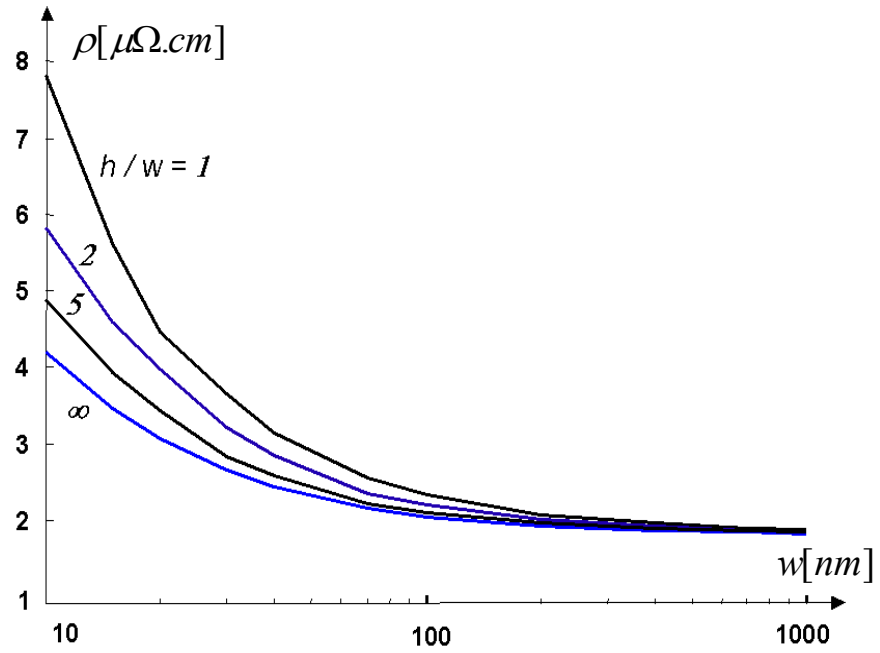



Figure 2.7: Resistivity of a thin wire versus wire width ($w$) for wires with different aspect ratios ($h/w$). Curves are for $p$ = 0 and at zero frequency

$$\frac{\sigma_0}{\sigma} = 1 + (1-p)\frac{3}{16}\frac{\lambda P}{A}. \tag{2.33}$$

## 2.5 Anomalous Skin Effect (ASE)

At sufficiently low temperatures and high frequencies, the mean free path of the electrons in a good conductor becomes greater than the classically predicted skin depth, and the classical skin effect equations break down. The same approach can be used to find the effect of both surface scattering and the skin effect on the resistivity of a thin film. If we assume that $\omega\tau$ is much smaller than one (where $\omega$ is frequency of the input signal and $\tau$ is the relaxation time, at room temperature $\tau$=2.7×10$^{-14}$sec and for f=10GHz, $\omega\tau$ will be 1.6×10$^{-2}$), then the BTE in (2.15) for the thin film shown in Figure 2.4 will be simplified to

$$\frac{f_1}{\tau} + v_z\frac{\partial f_1}{\partial z} = \frac{eE_x(z)}{m^*}\frac{\partial f_0}{\partial v_x}. \tag{2.34}$$

In this case, the electric field is a function of $z$ and is the solution of Maxwell's equations at a single frequency

$$\begin{aligned} \vec{\nabla}\times\vec{E} &= -j\omega\mu\vec{H} \\ \vec{\nabla}\times\vec{H} &= \vec{J} + j\omega\varepsilon\vec{E} \end{aligned}. \tag{2.35}$$

Equation (2.35) for the thin film shown in Figure 2.4 will reduce to

$$\begin{aligned} \frac{dE_x(z)}{dz} &= -j\omega\mu H_y(z) \\ \frac{dH_y(z)}{dz} &= J_x(z) + j\omega\varepsilon E_x(z) \end{aligned}, \tag{2.36}$$

or

$$\frac{d^2E_x(z)}{dz^2} = -j\omega\mu J_x(z) + \omega^2\mu\varepsilon E_x(z). \tag{2.37}$$

To determine the ASE, (2.34), (2.37), and (2.16) should be solved simultaneously. First assume we are solving for the normal skin effect problem. Replacing $E_x(z) = \sigma J_x(z)$ in (2.37) will yield

$$J(z) = \tfrac{1}{2}\frac{E_0}{\sigma}(e^{-\alpha z} + e^{-\alpha(d-z)}), \tag{2.38}$$

where $\alpha$ is the reciprocal of skin depth

$$\alpha = \frac{1}{\delta} = \sqrt{\pi f \mu \sigma}\,. \tag{2.39}$$

For now, assume that the electric field with surface scattering remains unchanged; therefore

$$E_x(z) = \tfrac{1}{2}E_0(e^{-\alpha z} + e^{-\alpha(d-z)}), \tag{2.40}$$

and the solution of (2.34) by incorporating (2.38) will be

$$f_1(\vec{v}, z) = \frac{e\tau_0 E}{m^*}\frac{\partial f_0}{\partial v_x}\frac{1}{1-\alpha\tau_0 v_z}\left[e^{-\alpha z} + F(\vec{v})e^{-z/\tau_0 v_z}\right]. \tag{2.41}$$

Assuming two distribution functions, $f^+$ for electrons with $v_z > 0$ and $f^-$ for electrons with $v_z < 0$, and considering the specularity parameter $p$ at surfaces $z$=0 and $z$=$d$, the boundary conditions for (2.41) will be

$$\begin{cases} f_1^+(\vec{v}, z=0) = pf_1^-(-\vec{v}, z=0) & \text{with} \quad v_z > 0 \\ f_1^-(\vec{v}, z=d) = pf_1^+(-\vec{v}, z=d) & \text{with} \quad v_z < 0 \end{cases}, \tag{2.42}$$

and the solution will be

$$\begin{cases} f_1^+(\vec{v}, z) = \dfrac{e\tau_0 E}{m^*}\dfrac{\partial f_0}{\partial v_x}\dfrac{1}{1-\alpha\tau_0 v_z}\left[e^{-\alpha z} + F^+(v_z)e^{-z/\tau_0 v_z}\right] & v_z > 0 \\ f_1^-(\vec{v}, z) = \dfrac{e\tau_0 E}{m^*}\dfrac{\partial f_0}{\partial v_x}\dfrac{1}{1-\alpha\tau_0 v_z}\left[e^{-\alpha z} + F^-(v_z)e^{-z/\tau_0 v_z}\right] & v_z < 0 \end{cases}, \tag{2.43}$$

where

$$F^{+}(v)=\frac{e^{d/\tau_0 v}}{\eta(v)(e^{2d/\tau_0 v}-p^2)}\Big[\,p(p\eta(v)-1)e^{-\alpha d}+(p-\eta(v))e^{d/\tau_0 v}\,\Big]$$

$$F^{-}(v)=\frac{e^{-d/\tau_0 v}}{\eta(v)(e^{-2d/\tau_0 v}-p^2)}\Big[\,p(p\eta(v)-1)e^{d/\tau_0 v}+(p-\eta(v))e^{-\alpha d}\,\Big], \qquad (2.44)$$

$$where \quad \eta(v)=\frac{1+\alpha\tau_0 v}{1-\alpha\tau_0 v}$$

Replacing (2.43) in (2.16) and introducing spherical coordinates ($v$, $\theta$, $\varphi$), similar to what we did in the previous section, will yield

$$J(z)=-\frac{2\pi e^2 m^{*2}\tau v_F^3 E_0}{h^3}\int_0^{\pi/2}\sin^3\theta\left\{\frac{e^{-\alpha z}+F^{+}(v)e^{-z/\lambda\cos\theta}}{1-\alpha\lambda\cos\theta}+\frac{e^{-\alpha z}+F^{-}(-v)e^{z/\lambda\cos\theta}}{1+\alpha\lambda\cos\theta}\right\}d\theta, \quad (2.45)$$

which for $f=0$ ($\alpha=0$) reduces to (2.27). Equation (2.45) is the main contribution of this work on ASE modeling. The "Moment method" is used to solve (2.34), (2.37), and (2.16) simultaneously. The dependency of the electric field on distance from the surfaces in its general form can be estimated by the summation of exponential functions (Fourier expansion). The numerical iteration steps are as follows:

STEP 1 – Assume $E(z)$ as (2.40).

STEP 2 – Find $J(z)$ using (2.45).

STEP 3 – Replace $J(z)$ in (2.37) find $E(z)$.

STEP 4 – Use Fourier expansion as $E(z)=\sum c_n e^{j\frac{2\pi}{d}z}$.

STEP 5 – If the solution has not yet converged, go to STEP 2.

Figure 2.8 shows the results as the ratio of wire resistivity to bulk resistivity. Calculations have been modified for a wire with a square cross-section, excited with a 10 GHz signal. These results show that when the wire dimension is below the skin depth at 10 GHz (≅650 nm) only *surface scattering* causes the increase in resistivity, while for the wires larger than the skin depth, both the skin effect and surface scattering are effective. For a 1 μm wire with a square cross-section, the *anomalous skin effect* increases wire

resistivity up to 40%, while the conventional skin effect shows only a 4% increase in the resistivity.

As a rule of thumb, the resistivity of a wire can be considered its dc resistivity up to the corner frequency $f_c$, where $f_c$ is the frequency at which the skin depth equals the wire width (sometimes 1/3 of wire width). Above $f_c$, wire resistivity depends on the square root of the frequency ($\sqrt{f}$). In the case of ASE we have two corner frequencies, $f_c$ and $f'_c$, where $f'_c$ is the frequency at which the skin depth is equal to five times the mean free path of electrons. Wire resistivity below $f_c$ is equal to the dc resistivity, while between $f_c$ and $f'_c$ it depends on the square root of frequency ($\sqrt{f}$). Above $f_c$ depends on frequency to the power of 2/3 ($\sqrt[3]{f^2}$). For copper wires and at room temperature $f'_c \cong 200$ GHz and

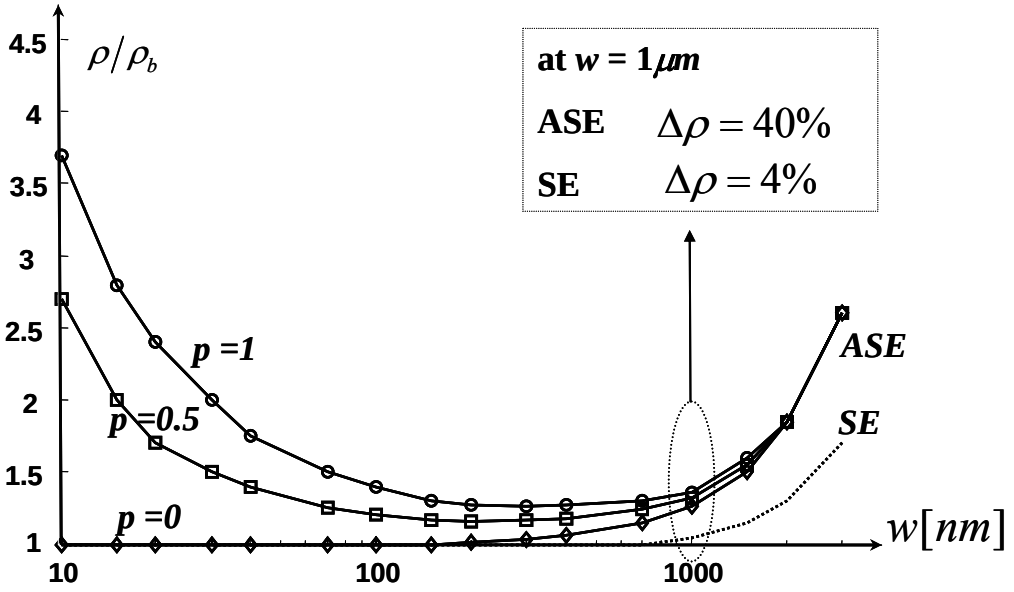



Figure 2.8: Ratio of wire resistivity to bulk resistivity for ASE (at $f$=10GHz) of a thin wire with aspect ratio of one ($w = h$) versus wire width, for different values of $p$; comparing to results for skin effect (SE).

at 100K it will be as low as $f'_c \cong 1$ GHz. Throughout this work, we remain at room temperature and frequencies below 100 GHz; therefore for delay and bit-rate limit calculations, we will assume that $\rho \propto \sqrt{f}$ . Analysis for frequencies above $f'_c$ needs to be done in the future work.

## 2.6 Surface Roughness and Barrier Thickness

In the past, aluminum was the main interconnect material. However, as dimentions shrink to sub-micron dimensions, the *rc* delay problem of interconnects appears. The sub-micron aluminum interconnect cannot handle high power density applications because of its relatively low resistance to electromigration. Because of these problems, the semiconductor industry focused on replacing aluminum with other metals. Copper seemed to be the best candidate because of its lower resistivity (about two thirds that of aluminum) and high electromigration resistance. Cost reduction was another reason. Copper techniques require 20% to 30% fewer steps than conventional aluminum patterning due to the new damascene approach and the higher packing density of the smaller feature size.

One of the challenges for Cu chemical mechanical polishing (CMP) is to control the uniformity of the surface topography while the interconnect layers increase to meet the more stringent die-level planarity requirement. The copper damascene technique circumvents metal etching, using the CMP process to form copper wiring inside the trenched oxide. The CMP process must be integrated horizontally and vertically to achieve high-quality process performance. Horizontal integration, which includes reliability of the consumables, development of new slurries, wafer cleaning systems, and a new metrology for endpoint detection, focuses on an integral solution for a robust yield and throughput for the CMP. Vertical integration integrates upstream processes such as copper barrier deposition and etching and downstream processes such as ILD deposition and lithography [2].

Copper diffuses into silicon oxide ($SiO_2$) and into silicon (Si) with disastrous effects so that barriers are required. Currently titanium nitride (TiN) and tantalum (Ta) are used. The ITRS predicts that barrier thickness will remain about 15% of the wire dimensions. Hence, in the semiconductor industry, copper resistivity is known as 2.2 μΩcm, instead of 1.7 μΩcm for bulk copper.

Surface roughness is described in different ways such as height of protuberances, aspect ratio, shape, distribution, and density of surface features and has been measured by different parameters such as RMS values, auto-covariance, and power spectrum. We will use the same model as in [36] to introduce surface roughness in terms of RMS values. Consider a cross-sectional view of a thin film as shown in Figure 2.9. Suppose that the roughness is approximated as a sine wave with sinusoidal undulations with wavelength $L$, of amplitude $h_1$ on the top surface and $h_2$ on the bottom surface. Assume there is a phase shift, $\varphi$, between these two. Hence the wire thickness can be written as

$$d(x) = d_0 + h_1 \sin(2\pi x / L) + h_2 \sin(2\pi x / L + \varphi). \tag{2.46}$$

$d(x)$ can be written as

$$d(x) = d_0 + h \sin(2\pi x / L), \tag{2.47}$$

where

$$h^2 = h_1^2 + h_2^2 + 2h_1 h_2 \cos(\varphi). \tag{2.48}$$

Using the same model for resistance calculations as in [36], the effective resistivity can be written as

$$\rho = \int_0^L \frac{\rho(x)}{d(x)} dx. \tag{2.49}$$

Obviously (2.49) is valid only when the wavelength of undulations is big enough! Using (2.33) for a thin film,

$$\rho(x) = \rho_0 (1 + \tfrac{3}{4}(1-p)\frac{\lambda}{d(x)}). \tag{2.50}$$

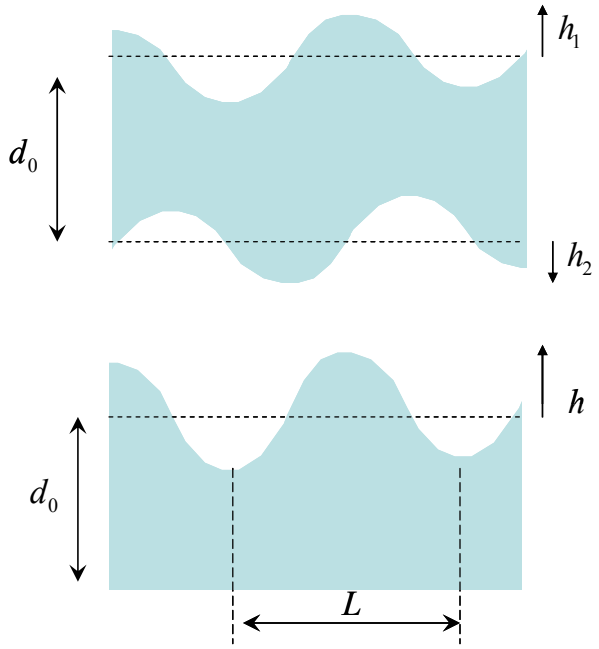



Figure 2.9: Cross-sectional view of a thin film with line edge roughness

By replacing (2.50) in (2.49) and assuming $d(x)$ as in (2.47), we obtain

$$\rho = \frac{\rho_0}{\sqrt{1-(h/d_0)^2}}\left[1+\frac{\frac{3}{4}(1-p)\lambda}{d_0-(h^2/d_0)}\right]. \tag{2.51}$$

This model was applied to a statistical Monte Carlo simulation framework to study die-to-die and with-in-die interconnect variations on the maximum critical path delay distribution. In Figure 2.10 (courtesy of [37]), the presented resistivity formula for surface roughness is compared to numerical simulations [38] and agrees well for a range of $w_0$ and $u$. Here, for the line-edge roughness (LER), $w_0$ is the effective line width and $u$ is the amplitude. For interconnect height undulation, $h_0$ is the nominal wire height and $v$ is half the dishing to occur after CMP.

Results indicate that effective line-width variations dominate LER variations until 2016, corresponding to a half-pitch of 22 nm. However, from 2016 to 2020, LER amplitudes will start to become a substantial percentage of the nominal effective line-width dimension, leading to an increase in the conductor resistivity. This effect amplifies the size-effects from traditional interconnect scaling, resulting in an increase in the maximum critical path delay mean and standard deviation. Since interconnect resistance is directly proportional to wire length, the 62% reduction in average wire length for a multi-core system compared to a single-core system maps to a ~35% reduction in the maximum critical path delay mean degradation and standard deviation for the year 2020 with a 14 nm half-pitch. As a result, multi-core systems are more tolerant to interconnect process variations and size-effects relative to single-core systems, leading to a more efficient design in performance, power, yield, and cost [37].

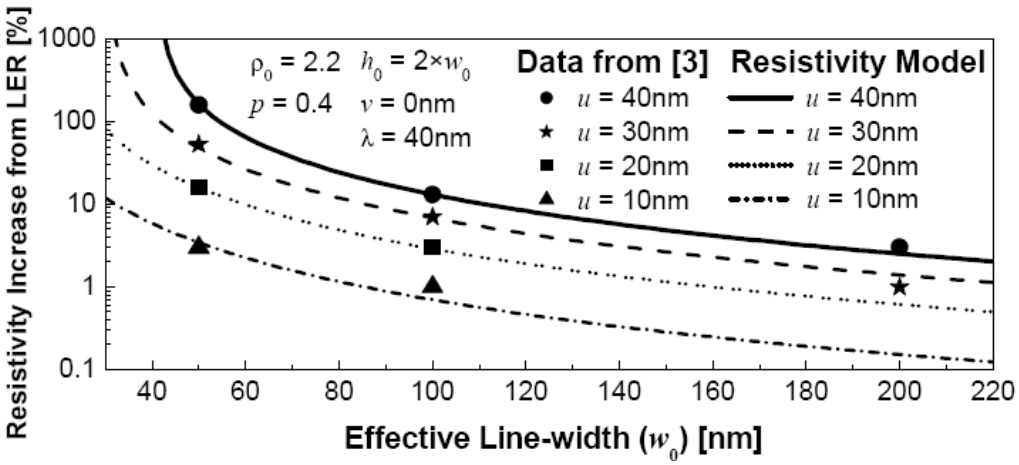



Figure 2.10: (courtesy of [37]) Comparison of resistivity model projections from our formula with numerical simulations from [38] for a range of $w_0$ and LER.

## 2.7 Conclusions

A new physical model for the anomalous skin effect is based on the Sondheimer approach and uses the moment method. Effective resistivity of thin wires with a rectangular cross-section as a function of wire dimensions and frequency is extracted. This model simplifies to the known asymptotic solutions for surface scattering and ASE at very high frequencies and applies to the region between the classical and anomalous skin effect. A compact formula for the impact of line-edge roughness on the resistivity of copper wires is presented.

# CHAPTER 3

# Delay and Bit-Rate Limit of Electrical Interconnects Using Phasor Analysis

## 3.1 Introduction

Electrical connections are part of a hierarchy that connects individual active devices at the lowest level to system-level connections at the highest [39, 40]. This hierarchy includes on-chip interconnects (consisting of local, semi-global, and global wires), multi-chip module interconnects, printed circuit board interconnects, and backplane (package) interconnects. The size of the interconnects varies with interconnect level both in cross-section and in length, at the chip level. The performance of many high-speed digital circuits is dominated by interconnects. The lumped *rc* delay model which has been previously used for many on-chip interconnects, is not adequate for GSI interconnects for several reasons: (1) clock frequencies of gigahertz and above, (2) faster gates and therefore sharper rise time, and (3) unscaleability of the length of global interconnects with technology [41].

Transmission line models are required for computer-aided design. The term "transmission line" in this respect means that the time delay along the line is important, meaning that the line inductance is included in the model. In the case of nearly all high bit-rate circuits, the interconnections between circuits require detailed analysis and modeling [42]. This includes routing on printed circuit boards and assessing the effect of bond wires and on-chip interconnects. However, not all on-chip interconnects are of equal importance to the performance of the IC. *Local* wires connect gates and cells within blocks. These wires get shorter with scaling and since the total resistance is dominated by

the resistance of the driver, the local wire's resistance does not matter much [42]. *Global* wires connect different blocks together and do not scale down when devices get smaller. As the driver resistance is smaller than the global wire resistance and since global wires are routed in top metal levels with big cross-sections, transmission line models need to be considered.

In this chapter, after reviewing transmission line theory, we show that for GSI applications, a quasi-TEM approximation is valid up to about 2 THz. Therefore, every on-chip interconnect in its most general form can be considered as a frequency-dependent distributed circuit that can be modeled by using the phasor analysis technique. This technique will be used to develop a new formula for the bit-rate limit of interconnects. Finally, the bit-rate limit of metal-clad polymer pins is studied.

## 3.2 Transmission Line as a Distributed Circuit

A transmission line is an electromagnetic guiding system for the efficient point-to-point transmission of electronic signals (information) and power. Transmission lines have been used in different types of electronic systems with a wide range of frequencies and applications. For on-chip interconnect modeling, in the past, *rc* tree models have been generally used. The signal delay through *rc* trees was often estimated using a form of the *Elmore delay* [43], which provided a dominant time constant approximation for monotonic step responses. At relatively higher signal speeds, the electrical length of the interconnect becomes a significant fraction of the operating wavelength. Consequently, the conventional lumped impedance interconnect models become inadequate, and transmission line models based on a quasi-transverse electromagnetic model (TEM) approximation are needed.

### 3.2.1 Transmission Line Model

The transmission line model can accurately describe the behavior of signal propagation along a line if the cross-section dimensions of the guiding structure are much smaller than the lowest characteristic wavelength of the signals propagating along the line. A distributed circuit model of the interconnections matches the circuit character of the overall electrical system [44, 45]. The inductance parameter *l* represents the series (loop) inductance per unit length of the line, and the capacitance parameter *c* represents the shunt capacitance per unit length between the two conductors. To represent line losses, the resistance parameter *r* is defined for the series resistance per unit length due to the finite conductivity of both conductors, while the conductance parameter *g* gives the shunt conductance per unit length of line due to dielectric loss in the material surrounding the conductors. The *r*, *l*, *g*, and *c* transmission line parameters of a configuration shown in Figure 3.1 can be derived in terms of the electric and magnetic field quantities by relating the corresponding energy stored and dissipated as:

$$l = \frac{\mu}{|I|^2} \int_S \mathbf{H}.\mathbf{H}^* ds \quad [H/m] \tag{3.1}$$

$$c = \frac{\varepsilon}{|V|^2} \int_S \mathbf{E}.\mathbf{E}^* ds \quad [F/m] \tag{3.2}$$

$$r = \frac{r_s}{|I|^2} \oint_{C_1+C_2} \mathbf{H}.\mathbf{H}^* dl \quad [\Omega/m] \tag{3.3}$$

$$g = \frac{\omega\varepsilon \tan\delta}{|V|^2} \int_S \mathbf{E}.\mathbf{E}^* ds \quad [S/m] \tag{3.4}$$

where **E** and **H** are the electric and magnetic field vectors in phasor form, and "*" denotes the complex conjugate operation, $r_s$ ($=\sqrt{\pi\mu\rho f}$) is the surface resistance of the conductors, $\varepsilon$ is the permittivity and $\tan\delta$ is the loss tangent of the dielectric material surrounding the conductors, $C_1$+$C_2$, the line integration in (3.3), is along the contours

enclosing the two conductors, and *S* is the area enclosed by these two contours (all cross-sectional area outside two conductors). The electric and magnetic fields are described by Maxwell's equations:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \tag{3.5}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \varepsilon \frac{\partial \mathbf{E}}{\partial t} \tag{3.6}$$

#### 3.2.1.1 Ideal Lossless Transmission Line

We shall assume that the line is *ideal* if (a) the two parallel conductors, which are made of arbitrary cross-sections with parallel generators (Figure 3.1), are made of a fictitious metal that is a perfect conductor ($\rho \to 0$); (b) the unbounded medium between the conductors is a vacuum, or a *perfect* dielectric. A perfect dielectric has a relative permittivity $\varepsilon_r (= \varepsilon / \varepsilon_0)$ and a relative permeability $\mu_r (= \mu / \mu_0)$, neither of which needs to be unity, but

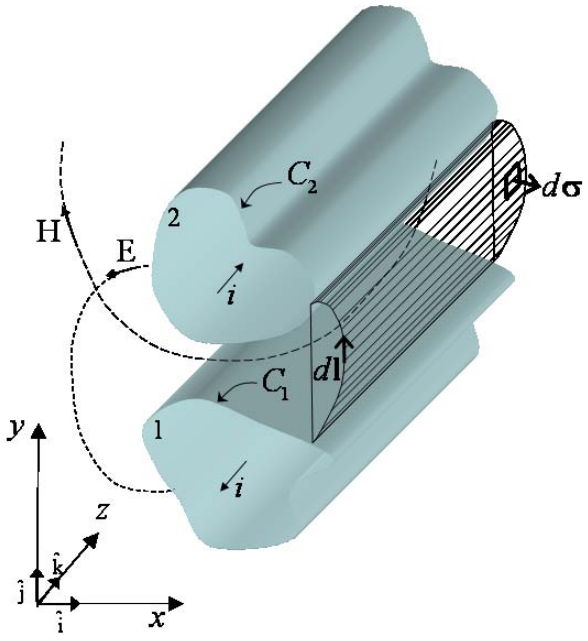



Figure 3.1: Physical structure of a transmission line.

which produces no dielectric or magnetic loss. Hence for a lossless transmission line, the boundary conditions imposed by Maxwell's equations will be as **E** is perpendicular to the surface of the conductor and **H** is tangential to the surface of the conductor which means that **E** and **H** have no component in the *z* direction (TEM mode) [46]. From the *z* component in equation (3.5) (i.e. $H_z = 0$), the following expression is obtained for the electric field:

$$\nabla \times \mathbf{E}_t = (\frac{\partial E_y}{\partial x} + \frac{\partial E_x}{\partial y})\hat{\mathbf{k}} = 0 \tag{3.7}$$

which means there is an equivalent for the electrostatic potential in the plane orthogonal to the direction of propagation of the wave for each value of *z* that can be written as

$$v(z) = \int_1^2 \mathbf{E}_t \cdot d\mathbf{l} \quad [V] \tag{3.8}$$

where the path of integration is arbitrary between conductors 1 and 2 as shown in Figure 3.1 and $\mathbf{l}$ is the unit vector along the path. Equation (3.5) in the *x* and *y* direction will result in

$$\frac{\partial E_x}{\partial z} = -\mu \frac{\partial H_y}{\partial t} \tag{3.9}$$

$$\frac{\partial E_y}{\partial z} = \mu \frac{\partial H_x}{\partial t} \tag{3.10}$$

From (3.8), (3.9), and (3.10),

$$\begin{aligned} \frac{\partial v}{\partial t} = -\int_1^2 \frac{\partial E_x}{\partial z} dx + \frac{\partial E_y}{\partial z} dy = -\int_1^2 -\frac{\partial B_y}{\partial t} dx + \frac{\partial B_x}{\partial t} dy = \\ -\frac{\partial}{\partial t}\int_1^2 -B_y dx + B_x dy = -\frac{\partial}{\partial t}\int_1^2 \mathbf{B} \cdot d\boldsymbol{\sigma} \end{aligned} \tag{3.11}$$

where, $d\boldsymbol{\sigma} = \hat{\mathbf{k}} \times d\mathbf{l}$ as shown in Figure 3.1. As $\mathbf{B} = \mu\mathbf{H}$ using the definition of inductance as in (3.1), (3.11) can be written as the first Telegrapher's equation:

$$\frac{\partial v}{\partial z} = -l \frac{\partial i}{\partial t} \tag{3.12}$$

where $i$ is the current in any one of the conductors in the direction of propagation.

Similarly, the current in one of the conductors can be obtained from the usual definition of magnetostatics. Assuming that current density is only non-zero inside the conductor and it is only in the direction of propagation, (3.6) will be written as

$$\nabla \times \mathbf{H}_t = (\frac{\partial H_y}{\partial x} + \frac{\partial H_x}{\partial y})\hat{\mathbf{k}} = J_z\hat{\mathbf{k}} \tag{3.13}$$

$$\frac{\partial H_x}{\partial z} = \varepsilon \frac{\partial E_y}{\partial t} \tag{3.14}$$

$$\frac{\partial H_y}{\partial z} = -\varepsilon \frac{\partial E_x}{\partial t} \tag{3.15}$$

Therefore,

$$i = \oint \mathbf{H}_t \cdot d\mathbf{l}\,. \tag{3.16}$$

And since, $\mathbf{D} = \varepsilon\mathbf{E}$

$$\frac{\partial i}{\partial z} = -\oint \frac{\partial D_y}{\partial t}dx - \frac{\partial D_x}{\partial t}dy = -\frac{\partial}{\partial t}\oint -D_y dx + D_x dy = -\frac{\partial}{\partial t}\int \mathbf{D}\cdot d\boldsymbol{\sigma} \tag{3.17}$$

Using the definition of capacitance per unit length as (3.2), the second Telegrapher's equation is obtained:

$$\frac{\partial i}{\partial z} = -c\frac{\partial v}{\partial t} \tag{3.18}$$

Equations (3.12) and (3.18) can be represented by a distributed $lc$ circuit, as shown in Figure 3.2. These equations represent a wave equation for voltage and current with the propagation constant given by $\gamma$. With a phasor formulation,

$$\begin{aligned} v &= \mathrm{V}(z)e^{j\omega t} \\ i &= \mathrm{I}(z)e^{j\omega t} \end{aligned} \tag{3.19}$$

with $\gamma^2 = -lc\omega^2$. The solution is

$$\begin{aligned} \mathrm{V}(z) &= \mathrm{V}_0^+ e^{-j\gamma z} + \mathrm{V}_0^- e^{+j\gamma z} \\ \mathrm{I}(z) &= \mathrm{I}_0^+ e^{-j\gamma z} + \mathrm{I}_0^- e^{+j\gamma z} \end{aligned} \tag{3.20}$$

Using either (3.12) or (3.18), the relation between current and voltage amplitudes will be:

$$V_0^{\pm} = \pm Z_0 I_0^{\pm}, \tag{3.21}$$

where $Z_0$ is defined as *characteristic impedance*:

$$Z_0 = \sqrt{\frac{l}{c}} \quad [\Omega] \tag{3.22}$$

On the other hand, $\gamma$ gives the velocity of propagation:

$$v_p = \frac{\omega}{\gamma} = \frac{1}{\sqrt{lc}} \quad [m/s] \tag{3.23}$$

#### 3.2.1.2 Lossy Transmission Line

Real conductors and dielectrics will show losses because of finite conductivity and resistivity, respectively. In this case, the conditions enumerated in the previous section are no longer valid. However, if losses are small, it is still possible to find an approximate interpretation in terms of quasi-static voltage and current in the orthogonal plane. As it is

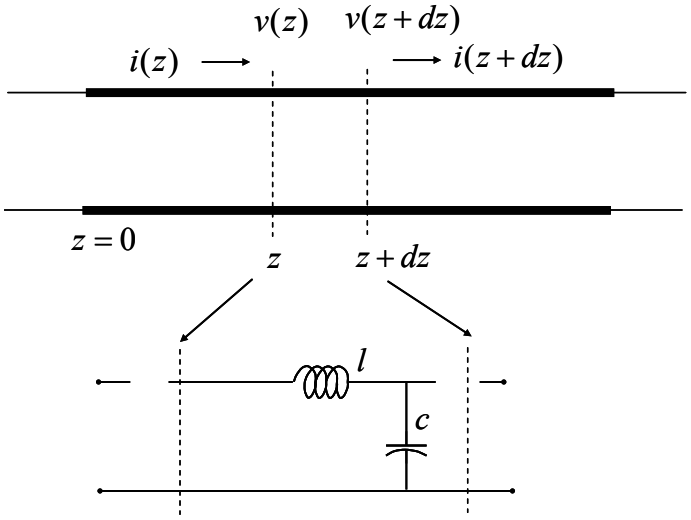



Figure 3.2: Distributed *lc* representation of Telegrapher's equation for a lossless transmission line.

obtained from a perturbation of the solution for the TEM model, it is called the quasi-TEM model and the Telegrapher's equation will be

$$\frac{\partial v}{\partial z} = -ri - l\frac{\partial i}{\partial t} \tag{3.24}$$

$$\frac{\partial i}{\partial z} = -gv - c\frac{\partial v}{\partial t} \tag{3.25}$$

where $r$ $[\Omega/m]$ and $g$ $[S/m]$ are per-unit-length longitudinal resistance and transverse conductance of the line, respectively. The values of $r$ and $g$ are obtained from a static analysis of the cross-section; therefore they are independent of frequency.

The parameter $r$ represents the per-unit-length voltage drop resulting from the conductor ohmic losses when a constant unitary current is flowing in one conductor and returning via the other. The parameter $g$ represents the per-unit-length transverse current due to the embedded medium losses when a constant unitary voltage is applied between the two conductors. In fact, $\frac{1}{2}(ri^2 + gv^2)$ represents the per-unit-length power dissipated along the guiding structure.

The following important properties hold for transmission lines surrounded by homogeneous media:

$$lc = \mu\varepsilon \tag{3.26}$$

$$gl = \mu\sigma \tag{3.27}$$

where $\sigma$ is the electrical conductivity of the embedding medium.

The harmonic wave solution to (3.24) and (3.25) will have a phase velocity as

$$v_{ph} = c\frac{\omega}{\sqrt{\omega^2 + \nu^2}} \tag{3.28}$$

where $\omega$ is the phasor wave frequency and $\nu$ is

$$\nu = \frac{1}{2}\left(\frac{r}{l} - \frac{g}{c}\right) \quad [1/\text{sec}] \tag{3.29}$$

Assuming small losses ( $r \ll \omega l$ and $g \ll \omega c$ ), the expression for characteristic impedance is [47]

$$Z_0 = \sqrt{\frac{r + j\omega l}{g + j\omega c}} \,. \tag{3.30}$$

As phase velocity $v_{ph}$ depends upon the frequency, any non-sinusoidal signal is propagated with distortion. Therefore, a lossy two-conductor transmission line is dispersive for $\nu \neq 0$. A lossy transmission line will be dispersionless if and only if

$$\nu = 0 \,, that\, is, \quad \frac{r}{l} = \frac{g}{c} \tag{3.31}$$

which is known as the *Heaviside condition* and has been important for telegraphy [47].

#### 3.2.1.3 Transmission Line with Frequency-Dependent Parameters

Interconnections in today's processor chips can be as long as 30 mm, operating up to several GHz rates. Studies show that the *rlcg* model is not valid for high frequencies [48]. Among the most dominant corrections for high frequencies is that resistance and inductance (in the most general form) should be considered frequency-dependent components. In the most general from, in the absence of distributed sources, a transmission line can be modeled in the Laplace domain with impedance and admittance per-unit-length $z(s)$ and $y(s)$. The line characteristic impedance is

$$Z_0(s) = \sqrt{\frac{z(s)}{y(s)}} \,. \tag{3.32}$$

and the propagation operator is

$$\gamma(s) = e^{-\sqrt{z(s)y(s)}} \,. \tag{3.33}$$

The per-unit-length impedance and admittance functions depend on the electromagnetic behavior of the guiding structure modeled by the transmission line. Evaluated along the imaginary axis of the $s$-plane, $z(j\omega)$ and $y(j\omega)$ describe the electrical behavior of the

line in the frequency domain. In the quasi-TEM approximation there are two reasons why the line parameters depend on the frequency, and they are profoundly different in nature. The line parameters *r*, *l*, *c*, and *g* depend on both the time dispersion of the materials and the transverse distribution of the electromagnetic field. Because of losses in the dielectric and the conductors, the transverse distribution of the electromagnetic field varies as the frequency varies, and hence the parameters *r*, *l*, *c*, and *g* also vary [49].

When the quasi-TEM hypothesis is no longer satisfied, the transmission line model can still be applied. In an ideal guiding structure, the behavior of each higher propagation mode (TE or TM) can be described in terms of an equivalent transmission line whose parameters *l* and *c* are frequency dependent and can be obtained by solving Maxwell's equations. In the most general case of an open lossy interconnection with a nonhomogeneous dielectric, operating at high frequencies, an equivalent frequency-dependent transmission line model can still be found by using a full-wave approach [15, 50].

In general, the expression of $y(j\omega)$, and hence of $y(s)$, depends on the actual geometry and material properties of the guiding structure modeled by the line. However, for all cases of interest, the asymptotic expression of $y(s)$ is of the type

$$\gamma(s) = (sc_{\infty} + g) + y_r(s) \tag{3.34}$$

where the term $sc_{\infty} + g$ represents the loading term of the admittance for $s \to \infty$, and the remainder $y_r(s)$ at infinity may be expressed through a Taylor series in terms of $s^{-1}$ and hence has the following asymptotic behavior:

$$y_r(s) \approx O(s^{-1}) \quad for\ s \to \infty \tag{3.34}$$

In general, the behavior of per-unit-length impedance $z(j\omega)$ can be described by assuming that the transmission line is made by two identical conductors and that the distance between the conductors is much larger than their transverse dimensions. In this case the per-unit-length impedance may be expressed as

$$z(j\omega) = 2z_s(j\omega) + j\omega l_e \tag{3.35}$$

where $l_e$ is the per-unit-length external self-impedance and $z_s(j\omega)$ is the per-unit-length surface impedance of the single conductor [51]. At high frequencies the electromagnetic field and current density are confined at the surface to a layer of thickness equal to a few skin depths, which is introduced as

$$\delta(\omega) = \sqrt{2/\omega\mu\sigma} \tag{3.36}$$

In this case, an approximate expression for the surface impedance is given by

$$z_s(j\omega) = \tfrac{1}{2}(r_{dc} + K\sqrt{j\omega}) \tag{3.37}$$

where $r_{dc}/2$ is the per-unit-length dc longitudinal resistance of the single conductor and *K* is a constant depending on the geometry and the physical parameters of the conductors.

Although the actual frequency behavior of the per-unit-length impedance is much more complicated, for all cases of interest, the asymptotic expression of *z*(*s*) is of the type

$$z(s) = r_{dc} + K\sqrt{s} + sl_{\infty} + z_r(s) \tag{3.34}$$

where the term $r_{dc} + sl_{\infty} + K\sqrt{s}$ represents the loading term of the impedance for $s \to \infty$, and the remainder $z_r(s)$ at infinity may be expressed through a Laurent series in terms of $1/\sqrt{s}$ and hence has the following asymptotic behavior [47, 49]:

$$z_r(s) \approx O(s^{-1/2}) \quad for\ s \to \infty \tag{3.34}$$

### 3.2.2 Validity of Distributed Model

The quasi-TEM model was derived using several assumptions about the propagating fields. The finite conductivity will generate a longitudinal electric field due to Ohm's law. The skin effect will cause an increase in resistance and a decrease in inductance for high frequencies. All these effects may lead to deviations from pure TEM wave propagation and therefore make the electrical models derived from it invalid. In cases in which

the losses are very significant the quasi-TEM assumptions will be invalid and a full-wave analysis is necessary for analyzing signal transmission.

The analysis is based on the condition that the longitudinal electric and magnetic fields are much smaller than their transverse components, but on a spatial average rather than on a point-to-point basis [52]. Assuming that the conductivity of the surrounding material is negligible ( $g \approx 0$ ), the range of validity for the quasi-TEM model is

$$w \ll \frac{1}{\omega\sqrt{\mu_0 \langle \varepsilon(\mathrm{r}) \rangle}} \tag{3.35}$$

where $w[m]$ is the cross-section of interconnects and $\langle \varepsilon(\mathrm{r}) \rangle$ is the spatial average of permittivity. For VLSI applications, $\langle \varepsilon(\mathrm{r}) \rangle$ is close to the permittivity of $SiO_2$ ≈ 35.4 pF/m. If we take 10 μm as a typical maximum value for the cross-sectional dimensions of interconnects, the validity of the quasi-TEM model holds for signal frequencies such that:

$$\omega \ll 1.5 \cdot 10^{13} \text{ rad/s} = 2.4 \text{ THz} \tag{3.36}$$

which is beyond significant frequencies of digital signals in integrated circuits. So, the quasi-TEM model is clearly applicable in these cases. In the previous chapter we showed that at room temperature anomalous skin effect will appear at frequencies around 200 GHz, therefore quasi-TEM assumptions are valid even for the anomalous skin effect region.

## 3.3 Phasor Analysis

Consider a transmission line with length $x$ excited with arbitrary input $v_s$ through a source impedance $Z_s$ and terminated with a load impedance $Z_l$, as shown in Figure 3.3. A general transmission line can be considered a distributed $zy$ (in contrast with distributed $rlc$), where $z(s)$ is the impedance per unit length and $y(s)$ is the admittance per unit length of the line as shown in Figure 3.4. The Telegrapher's equations will be

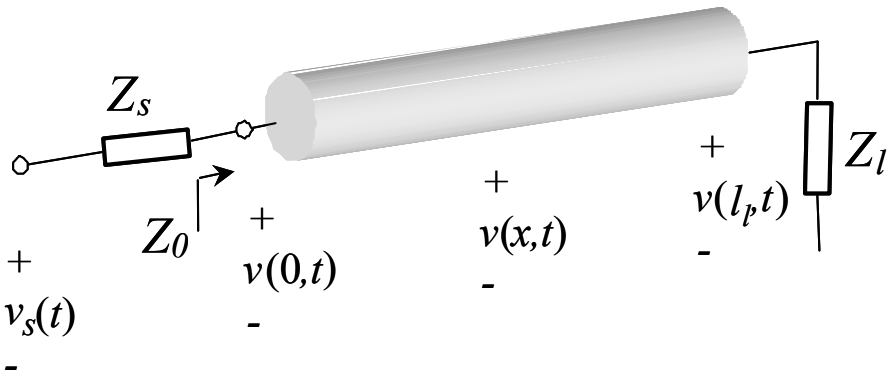



Figure 3.3: Transmission line structure used in "phasor analysis"

$$\frac{\partial v}{\partial x} = -z(s)i \tag{3.37}$$

$$\frac{\partial i}{\partial z} = -y(s)v \tag{3.38}$$

Assuming $v_0$ is the voltage at point $x$=0, the voltage that is traveling in the positive direction of $x$ can be written as

$$v(x,s) = v_0(s)e^{-\gamma(s)x} \tag{3.39}$$

Where the propagation constant $\gamma$ is

$$\gamma(s) = \sqrt{z(s)y(s)} \tag{3.40}$$

The phasor representation of the input signal can be written as

$$v_s(t) = \mathrm{Re}\left[\sum_{n=0}^{\infty} \mathbf{v}_n e^{jn\omega_0 t}\right] \tag{3.41}$$

while the voltage along the line at arbitrary position $x$ and time $t$ can be written as

$$v(x,t) = \mathrm{Re}\left[\sum_{n=0}^{\infty} \mathbf{V_n}(x) e^{jn\omega_0 t}\right] \tag{3.42}$$

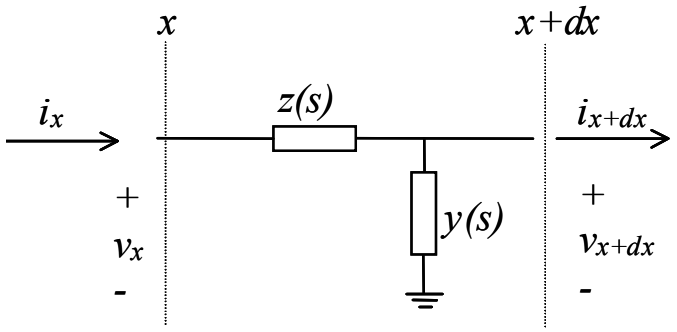



Figure 3.4: General equivalent circuit of a distributed transmission line.

For a semi-infinite line the phasor voltage at $x$=0 can be written as

$$\mathbf{V_n}(0) = k_{1n}\mathbf{v}_n \tag{3.43}$$

where $k_{1n}$ is

$$k_{1n} = \left.\frac{Z_0(s)}{Z(s)+Z_0(s)}\right|_{s=jn\omega_0} \tag{3.44}$$

and characteristic impedance of the line, $Z_0(s)$ is

$$Z_0(s) = \sqrt{z(s)/y(s)} \tag{3.45}$$

Using (3.39) the phasor voltage at $x$ will be written as

$$\mathbf{V}_n(x) = k_{1n}\mathbf{v}_n e^{-\gamma_n x} \tag{3.46}$$

where

$$\gamma_n = \gamma(s)\big|_{s=jn\omega_0} \tag{3.47}$$

For a finite line, the reflections at both ends should be considered. As shown in Figure 3.5 the phasor voltage at point $x$ will be

$$\begin{aligned}\mathbf{V}_n(x) = {} & k_{1n}\mathbf{v}_n e^{-\gamma_n x} + k_{1n}k_{3n}\mathbf{v}_n e^{-\gamma_n l}e^{-\gamma_n x'} + k_{1n}k_{2n}k_{3n}\mathbf{v}_n e^{-2\gamma_n l}e^{-\gamma_n x} \\ & + k_{1n}k_{2n}k_{3n}^2\mathbf{v}_n e^{-3\gamma_n l}e^{-\gamma_n x'} + k_{1n}k_{2n}^2k_{3n}^2\mathbf{v}_n e^{-3\gamma_n l}e^{-\gamma_n x'} + \cdots\end{aligned} \tag{3.48}$$

or simply

$$\mathbf{V}_n(x) = k_{1n}\mathbf{v}_n \frac{e^{-\gamma_n x} + k_{3n}e^{\gamma_n x}e^{-2\gamma_n l}}{1 - k_{2n}k_{3n}e^{-2\gamma_n l}} \tag{3.49}$$

as $k_{2n}$ and $k_{3n}$ are reflection coefficients at the source and load, as

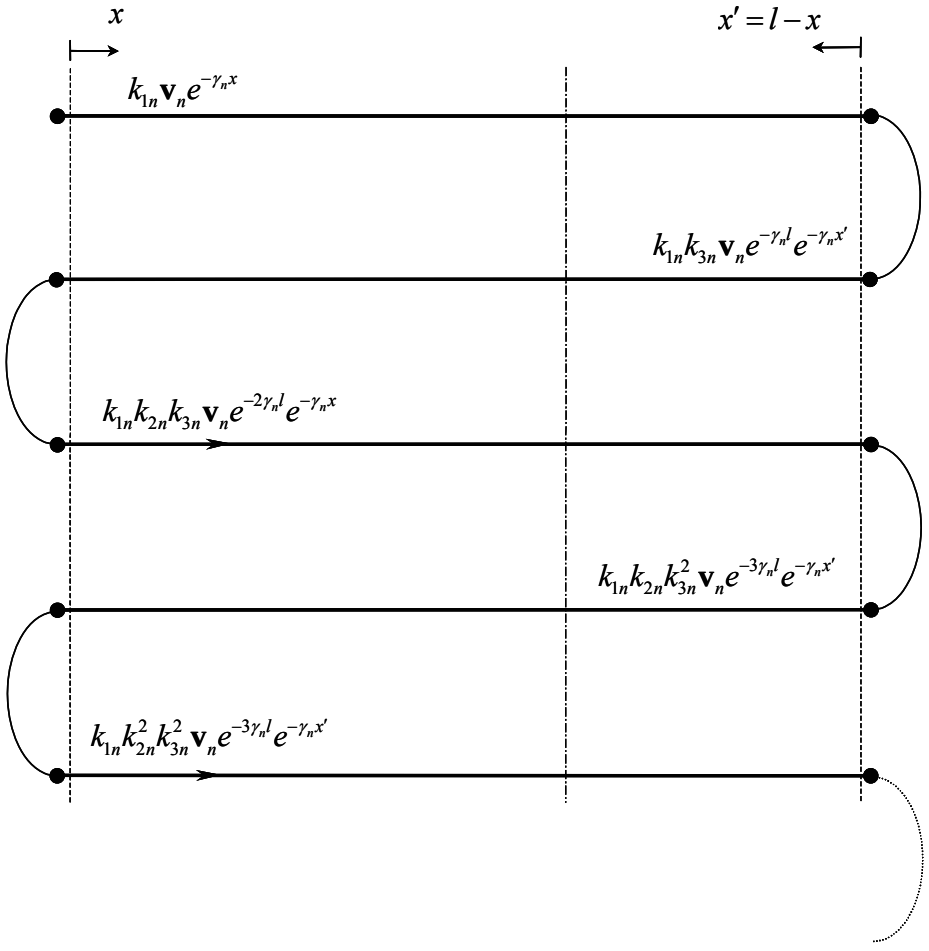



Figure 3.5: Reflections from source and load.

$$k_{2n} = \left. \frac{Z_s(s) - Z_0(s)}{Z_s(s) - Z_0(s)} \right|_{s=j\omega n} \tag{3.50}$$

$$k_{3n} = \left. \frac{Z_l(s) - Z_0(s)}{Z_l(s) - Z_0(s)} \right|_{s=j\omega n} \tag{3.51}$$

The proposed "phasor analysis" approach is validated by SPICE simulation. Two different test cases are considered. The results of the voltage at the beginning and at the end of the line are shown in Figure 3.6.

Test case 1 (Figure 3.6.a):

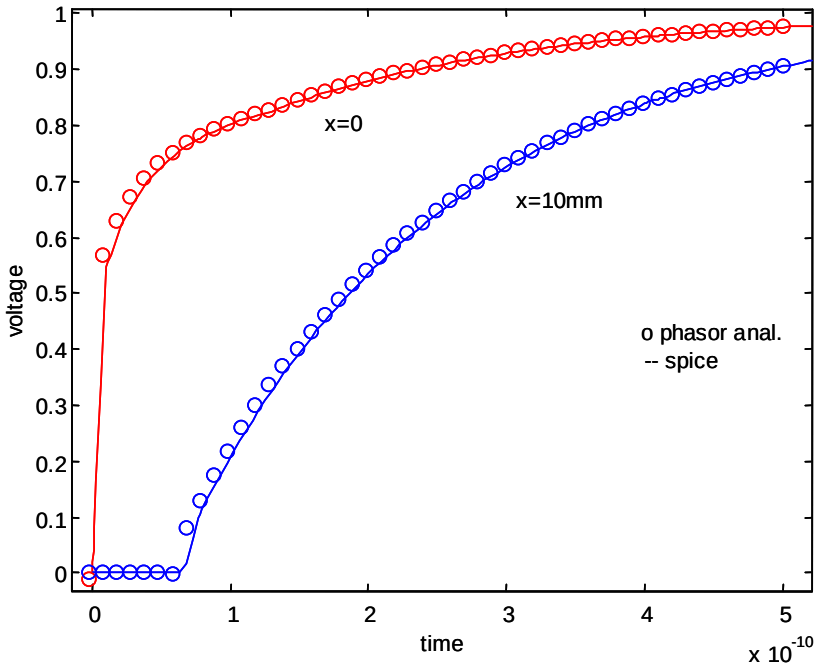
$z = r + sl$ , $y = g + sc$ , $r = 16.7\,[\Omega / mm]$, $\quad l = 2.1 \times 10^{-7}\,[H / mm]$, $\quad g = 0$

and $c = 2.1 \times 10^{-10}\,[F / mm]$. Total length of the line: $l_{line} = 10mm$

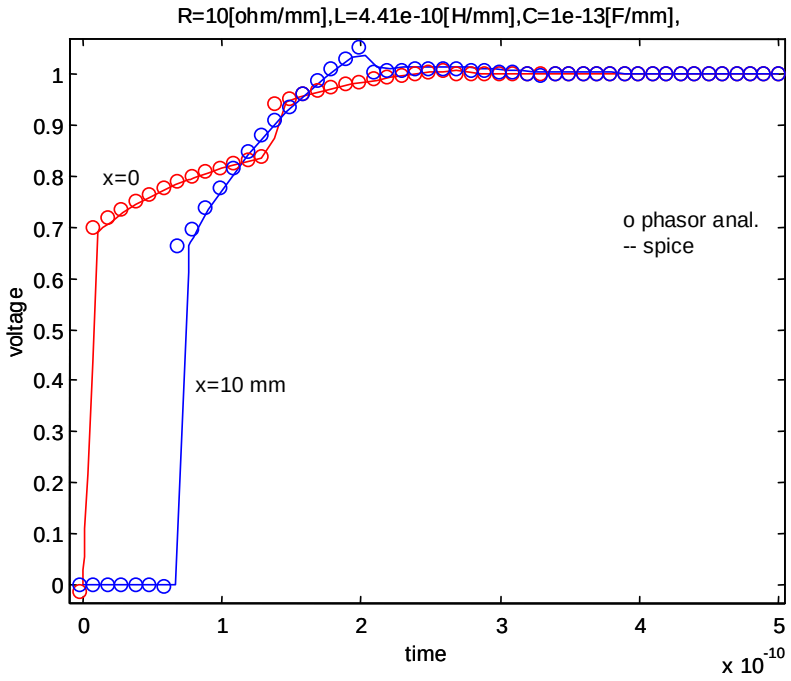
Test case 2 (Figure 3.6.b):

$z = r + sl$ , $y = g + sc$ , $r = 10\,[\Omega / mm]$, $\quad l = 4.41 \times 10^{-7}\,[H / mm]$, $\quad g = 0$

and $c = 1 \times 10^{-10}\,[F / mm]$. Total length of the line: $l_{line} = 10mm$




(a)




(b)

Figure 3.6: Comparing SPICE simulations (solid line) with results of phasor analysis (circles) for two different cases at $x$=0 and $x$=10 mm.

## 3.4 Delay Calculations of Electrical Interconnects Considering ASE

In this section the impact of ASE on the performance of a digital transmission line for the structure shown in Figure 3.7 is studied. The goal is to show how much the delay of the line changes for wires with different widths ($w$). By defining resistivity as in (2.32), the dependency of resistivity versus wire dimension and frequency of operation can be written as

$$\rho = \rho_b(\alpha_0 + \alpha_1\sqrt{f}) \tag{3.52}$$

where $f$ is the frequency of operation. The symbols $\alpha_0$ and $\alpha_1$ $[\sec^{0.5}]$ are coefficients which both depend on the wire dimension. *Surface scattering*, which increases resistivity at low frequencies, determines the dependency of $\alpha_0$ to the wire dimension, while the *anomalous skin effect* determines the value of $\alpha_1$. Table 3.1 shows $\alpha_0$ and $\alpha_1$ for different values of $d$ for the structure shown in Figure 3.7.

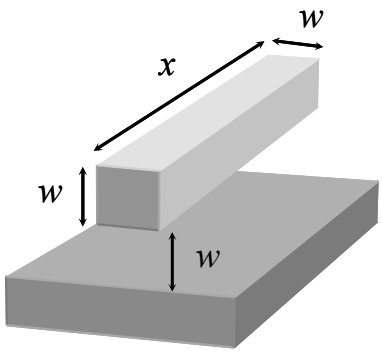



Figure 3.7: Structure that is chosen for the case study, wire with unity aspect ratio above a ground plane.

Table 3.1: Values of $\alpha_0$ and $\alpha_1$ for a copper line with $p = 0.5$. $x_{20G\text{-}ASE}$, $x_{20G\text{-}SE}$ and $x'_{20G}$ are the length at which the bit-rate limit of a wire is 20Gb/sec for ASE, skin effect and based on the low-loss approximation, respectively.

| $D$(nm) | $\alpha_0$ | $\alpha_1$ ($\sec^{0.5}$) | $x_{20G\text{-}ASE}$ (mm) | $x_{20G\text{-}SE}$ (mm) | $x'_{20G}$ (mm) |
|---|---|---|---|---|---|
| 10 | 2.64 | 1.4e-8 | 0.0077 | 0.0121 | 0.0231 |
| 20 | 1.71 | 1.8e-8 | 0.018 | 0.0232 | 0.0452 |
| 50 | 1.33 | 4e-7 | 0.048 | 0.0555 | 0.115 |
| 100 | 1.08 | 8e-7 | 0.104 | 0.109 | 0.231 |
| 500 | 1 | 3e-6 | 0.523 | 0.541 | 1.154 |
| 1000 | 1 | 6e-6 | 1.03 | 1.11 | 2.31 |
| 2000 | 1 | 1.2e-5 | 2.017 | 2.44 | 4.43 |
| 5000 | 1 | 3e-5 | 4.977 | 7.686 | 11.56 |
| 10000 | 1 | 6e-5 | 9.687 | 18.04 | 23.07 |

The delay of a line using bulk resistivity ($\rho_b$) is called $T_0$ and the delay considering the anomalous skin effect using (41) is called $T$. The percentage increase in the delay caused by the anomalous skin effect for a 1-mm long wire is shown in Figure 3.8. It is also compared with the relative increase in the delay by only considering the skin effect. It can be followed that for a line in the RC region (region 1), delay is mostly increased by the low-frequency component of the resistivity because of surface scattering. And for a line in the RLC region (region 3), as delay is almost time-of-flight (TOF), there is no change in the delay caused by resistivity increase, while at the region between the RC and RLC regions (region 2), the difference between the *surface scattering* and *skin effect* with the *ASE* is considerable. The *anomalous skin effect* shows a 20% increase in the delay of a 1-mm length of the wire, while this could be larger for longer wires (at $w$=500$nm$ in Figure 3.8). For a 1-mm interconnect, wires thinner than 100 nm are in the RC region and

those thicker than 1 μm are in the RLC region. The border between these regions depends on the length of the wire.

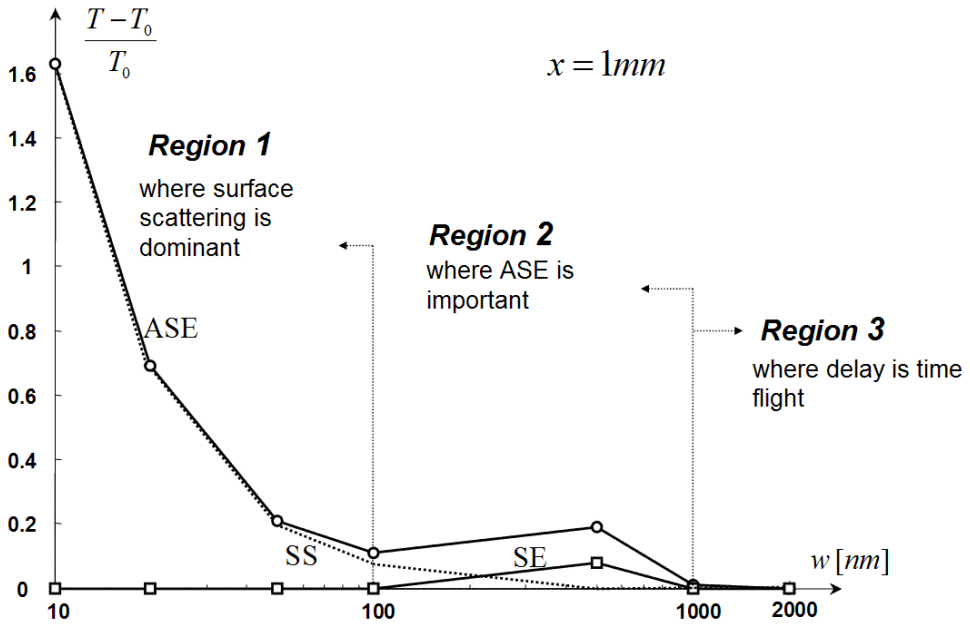



Figure 3.8: Comparison between relative increases in the delay of a 1 mm long wire versus wire width, caused by the anomalous skin effect and skin effect.

## 3.5 Bit-Rate Limit of Electrical Interconnects

The advance of modern IC processes has supported increasing bit rates in many applications, such as hard disk drives and optical networking. The reduced feature size of new generations of IC technology drives the improvement of the high-frequency performance of transistors and passive elements, but at the same time requires a reduction of supply voltages. This poses significant challenges to the design of high-frequency building blocks. Therefore, more than ever, it is necessary to have an accurate model for the bit-rate limit of electrical interconnects.

In this section, after introducing a normalization technique, previous bit-rate models (low-loss approximation) are studied. Compact models are presented for the bit-rate limit of transmission lines using a general form of resistance that for the first time simultaneously considers dc resistance, skin effect, and surface scattering. A conventional low-loss approximation that is only valid for fast rising signals is also relaxed. In contrast to previous models, it is shown that the bit-rate limit of a transmission line is not scale-invariant. It is also shown that the error of previous models is large if the bit-rate limit is not considerably larger than the reciprocal time-of-flight.

### 3.5.1 Normalization

In this section, parameters that are used to describe transmission lines are introduced. It is also shown how these parameters can be normalized to simplify equations. Using the argument that we made at the end of Section 2.5 and Section 3.2.1.3, the resistance per unit length of a wire in the Laplace domain can be written as

$$\mathbf{r}(s) = r + r'\sqrt{s} \quad [\Omega/m] \tag{3.53}$$

Hence, through this chapter, an $rr'lc$ line represents an $\mathbf{r}lc$, where $r$ and $r'$ relate to the resistance per unit length of a wire ($\mathbf{r}$), as shown in (3.53), and $l$ and $c$ are inductance and capacitance per unit length of the wire. Normalizing an $rr'lc$ line to a line with unity time-of-flight ($T_F$) and characteristic impedance makes the equations shorter and simpler, and makes the comparisons easier.

Suppose an $rr'lc$ line with length $x$ is driven by a source that has an impedance of $R_s+sL_s$ and is terminated by a load capacitance $C_l$. The concepts that are used for the normalization are as follows:

*i)* An $rr'lc$ line with length $x$, can be replaced by a $(xr)(xr')(xl)(xc)$ line with unity length.

*ii)* Multiplication of all impedances of any linear circuit by a certain value does not change the voltage-voltage transfer function of the system; therefore, all impedances can be multiplied by the reciprocal characteristic impedance of the line, $Z_0$.

*iii)* Finally, time-frequency duality implies that shrinking time by $T_F$ is the same as multiplying all frequencies by the same factor ($T_F$).

By using these three concepts, an $rr'lc$ line with length $x$ can be replaced by an $\tilde{r}\tilde{r}'11$ line with unity length. The following relations between timing in the original and normalized cases, however, should be considered:

$$\tilde{t} = t\frac{1}{T_F} \quad ; \quad \tilde{f} = fT_F \tag{3.54}$$

Throughout this section, all normalized parameters are marked with a tilda sign. Other parameters should also be normalized as

$$\tilde{R}_s = R_s\frac{1}{Z_0}\,;\ \tilde{L}_s = L_s\frac{1}{T_F Z_0}\,;\ \tilde{C}_l = C_l\frac{Z_0}{T_F}\,;\ \tilde{r} = r\frac{x}{Z_0}\,;\ \tilde{r}' = r'\frac{x}{\sqrt{T_F}Z_0} \tag{3.55}$$

where $Z_0 = \sqrt{l/c}$ is the characteristic impedance of the line.

### 3.5.2 Bit-rate Limit Model Neglecting DC Resistance with Low-Loss Approximation (Previous Models)

In this section, the bit-rate limit of an ideally terminated normalized line is calculated using the approximations used in previous models [16-18]. This brief review highlights the advantages of the new models presented in the next sections. The ideal matching is when a transmission line is terminated by a similar line with infinite length. Hence, assuming $\tilde{R}_s = \tilde{L}_s = 0$, the step response at the end of a normalized line is

$$v(s) = \frac{1}{s}\exp(-\sqrt{s(s + \tilde{r}'\sqrt{s} + \tilde{r})}) \tag{3.56}$$

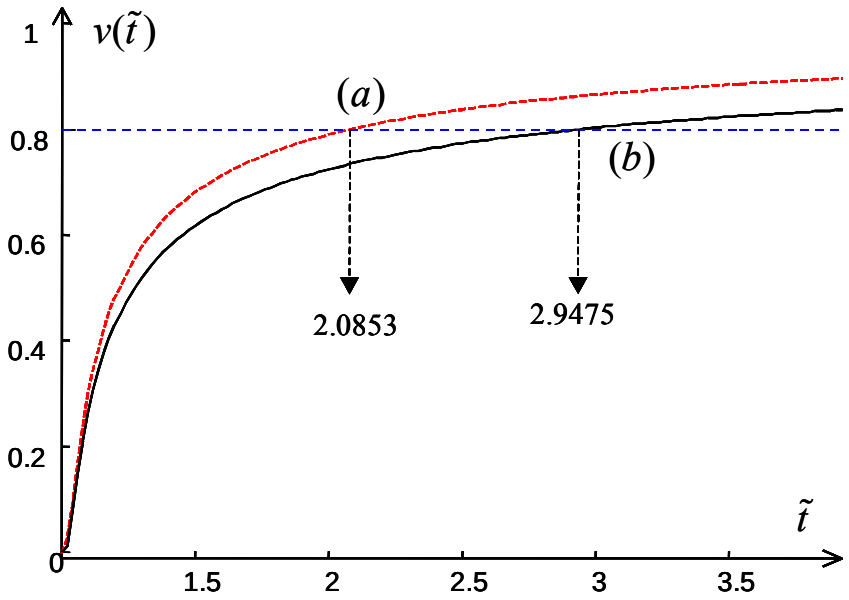



Figure 3.9: Step response in normalized time for $\tilde{r}'=1$ (*a*) is the exact solution of (3.57) and (*b*) is the conventional low-loss approximation as (3.60).

Ignoring the dc resistance, the step response in the normalized time domain can be written as

$$v(\tilde{t}) = L^{-1}\left\{\frac{1}{s}\exp(-\sqrt{s(s+\tilde{r}'\sqrt{s})})\right\} \tag{3.57}$$

Since the inverse Laplace transform of (3.56) is complicated, the following binomial approximation (low-loss approximation) is conventionally used:

$$\sqrt{s(s+\tilde{r}'\sqrt{s})} \approx s + \frac{1}{2}\tilde{r}'\sqrt{s} \tag{3.58}$$

which yields

$$v(\tilde{t}) = L^{-1}\left\{\frac{1}{s}e^{-s}e^{-\frac{1}{2}\tilde{r}'\sqrt{s}}\right\} \tag{3.59}$$

and therefore

$$v(\tilde{t}) = erfc\left(\frac{\tilde{r}'}{4\sqrt{\tilde{t}-1}}\right)u(\tilde{t}-1) \tag{3.60}$$

where *erfc*(.) is the complementary error function.

By knowing the step response, the eye opening can be written as $2s(T)$-1, where $s(t)$ is the step response and $T$ is the symbol time [17] (for instance, the symbol time $T_{0.8}$ is defined as $s(T_F+T_{0.8})=0.8$). Therefore, for an eye opening of 60%, $s(T)$ should be equal to 0.8. Solving (3.58), gives

$$\widetilde{B}_{60\%} = 1/T_{0.8} = 0.52\,\widetilde{r}'^{-2} \tag{3.61}$$

Rewriting the results of [16-18] in the normalized form yields a bit-rate formula that is the same as in (3.61). However, equation (3.61) underestimates the bit-rate limit of a wire. In Figure 3.9, the solution of (3.60) is compared with the exact solution of (3.57) when $\widetilde{r}'=1$. The approximation used in (3.60) predicts $T_{0.8}$=1.9475, while the actual solution of (3.57) is $T_{0.8}$=1.0853 (80% underestimation in the bit-rate limit).

### 3.5.3 Bit-Rate Limit Model Neglecting DC Resistance and Relaxing Low-Loss Approximation (New Model)

Previous models were based on the low-loss approximation used in (3.58). As shown in Figure 3.10, the error of using (3.61) is small if $\left|\widetilde{r}'/\sqrt{s}\right| << 1$, which is valid if only high-frequency components of the signal are important. This occurs when signal rise time is small. Therefore, for a line with a certain length, this approximation fails if the wire cross-sectional area is too small.

The approximation in (3.58) fails when $s$ is small (or when $\widetilde{t}$ is large). For the region that $\left|\widetilde{r}'/\sqrt{s}\right| >> 1$, (3.57) can be approximated as

$$v(\widetilde{t}) = L^{-1}\left\{\frac{1}{s}\exp(-\sqrt{s(s+\widetilde{r}'\sqrt{s})})\right\} \approx L^{-1}\left\{\frac{1}{s}e^{-\sqrt{\widetilde{r}'}s^{3/4}}\right\} \tag{3.62}$$

which yields

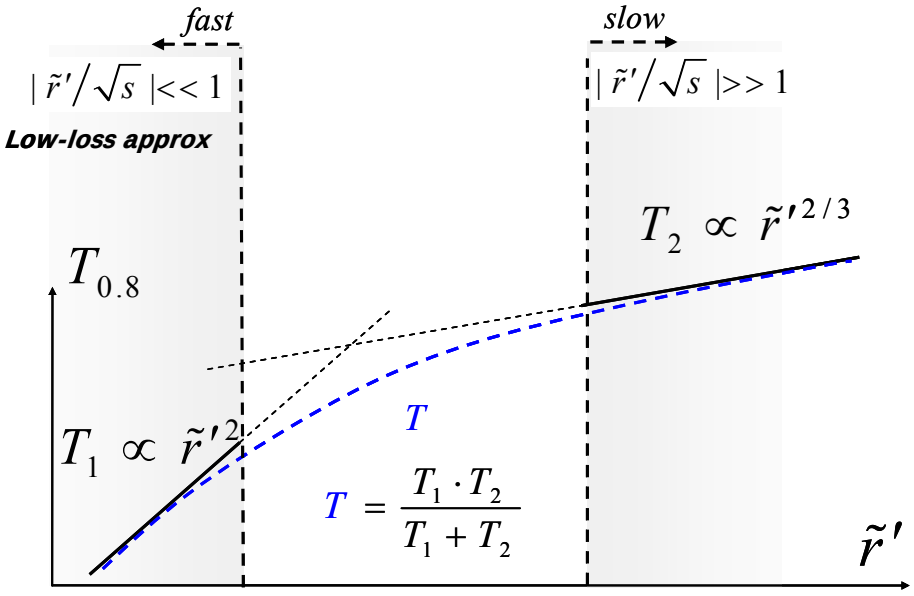



Figure 3.10: Two asymptotic solutions for fast and slow transmission lines.

$$v(\tilde{t}) = 1 - \frac{\sqrt{\tilde{r}'}}{2.8}\tilde{t}^{-3/4}. \tag{3.63}$$

By solving (3.63) for an eye opening of 60%, we have

$$\tilde{B}_{60\%} = 0.45\tilde{r}'^{-2/3} \tag{3.64}$$

Considering that (3.60) and (3.63) are the solutions for fast and slow lines, respectively, it can be heuristically deduced that the summation of reciprocal delays for these two cases will sum up into the delay that is valid for all regions. Therefore, the general formula for the bit rate is the summation of (3.61) and (3.64), and by de-normalizing the solution, the maximum bit rate of a wire for an eye opening of 60%, neglecting dc resistance, is found as

$$B_{60\%} = T_F^{-1}(0.52\tilde{r}'^{-2} + 0.45\tilde{r}'^{-2/3}) \tag{3.65}$$

In Figure 3.11, the normalized bit-rate limit is plotted versus $\tilde{r}'$. The differences between (3.61) and (3.65) show that the conventional bit-rate formulas –based on (3.61)– are just valid when the normalized bit-rate limit ($\tilde{B}_{60\%}$) is considerably larger than 1 (equal to bit-rate limit considerably greater than reciprocal $T_F$).

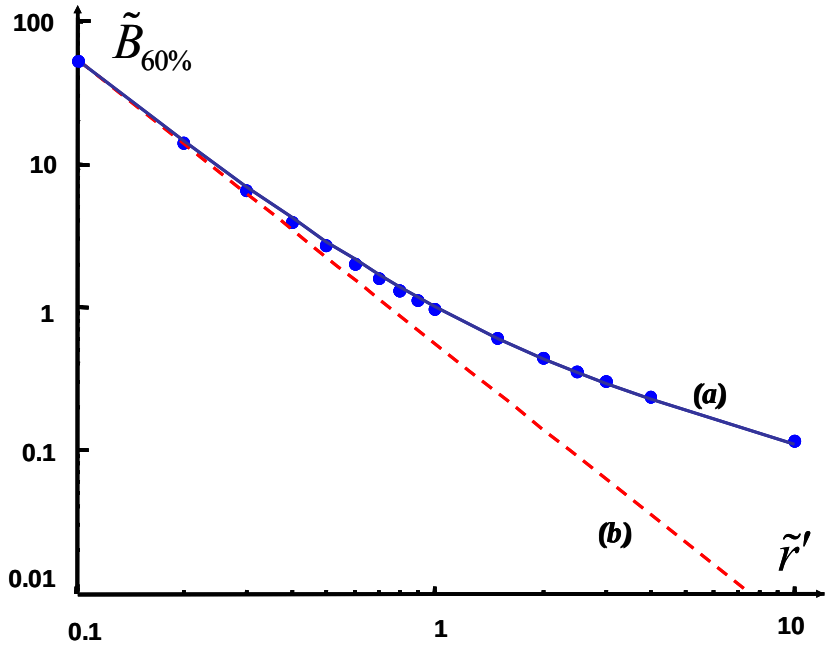



Figure 3.11: Normalized bit-rate limit vs. $\tilde{r}'$, neglecting dc resistance. (b) is the conventional scale-invariant relation shown in (3.61), (a) is the formula written in (3.65) and dot points are results of solving (3.57) numerically

### 3.5.4 Bit-Rate Limit Model Considering DC Resistance (New Model)

In all previous classical works on wire bit rate, the dc resistance of the wire has been neglected. In this section, the wire bit-rate limit is modeled, while the resistance per unit length of the wire is in its most general form as (3.53). Again, as was discussed in Section 3.5.2, for the bit-rate limit calculation, a matched load and ideal step input are considered.

The step response in the normalized notation is as (3.56). Two new variables, $B$ and $D$, are defined as

$$B = \frac{1}{2}(\frac{\tilde{r}'}{2})^2 \quad ; \quad D = (\frac{\tilde{r}'}{2})^2 - \tilde{r} \tag{3.66}$$

By using the Taylor series expansion of (3.56) around $s = \infty$, the step response in the Laplace domain can be written as

$$v(s) = e^{D/2} e^{-s} \sum_{n=0}^{\infty} a_n s^{-1-\frac{n}{2}} e^{-\sqrt{2Bs}} \tag{3.67}$$

where

$$\begin{aligned} a_0 &= 1 \\ a_1 &= -\frac{1}{2} D\sqrt{2B} \\ a_2 &= \frac{1}{8} D(D + 8B + 2BD) \\ a_3 &= -\frac{1}{48} D\sqrt{2B}(248D + 48B + 18D + 3D^2 + 2BD^2) \\ &\cdots \end{aligned} \tag{3.68}$$

If $D = 0$, which means

$$\tilde{r} = \left(r\tilde{'}/2\right)^2 \tag{3.69}$$

then all $a_i$'s (for $i \geq 1$ ) are equal to zero and (3.67) can be simplified to

$$v(s) = \frac{1}{s} e^{-s} e^{-\sqrt{2Bs}} \tag{3.70}$$

Therefore, the step response in the normalized time domain becomes equal to

$$v(\tilde{t}) = erfc\left(\frac{B}{2(\tilde{t} - 1)}\right) u(\tilde{t} - 1) = erfc\left(\frac{\tilde{r}'}{4\sqrt{\tilde{t} - 1}}\right) u(\tilde{t} - 1) \tag{3.71}$$

which is the same as (3.60). This shows that (3.60), which was the low-loss approximation (used in [16-18]) for the step response of a wire with zero dc resistance (and suffering from underestimation shown in Figure 3.10), is the exact solution of a wire when condition (3.69) is satisfied. For a wire with a rectangular cross-section the condition in (3.69) occurs when

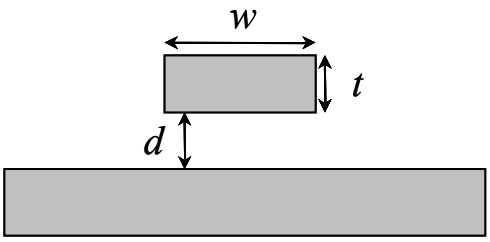



Figure 3.12: Wire above the ground plane structure that is chosen for the case study. $\varepsilon_r$ is chosen to be 3.9.

$$a_R = 2Z_0\sqrt{\varepsilon/\mu} \tag{3.72}$$

where $a_R = t/w$ is the aspect ratio of the wire. For the configuration shown in Figure 3.12, when $d$=$t$, (3.72) occurs around $a_R \approx 0.3$.

Again, finding the rigorous solution of (3.67) in the time domain is too complicated, but by knowing the solution of (3.67) in two cases ((3.61) and (3.65)), the bit-rate limit formula for any wire in its most general form is written as

$$B_{60\%} = T_F^{-1}(k_1\tilde{r}'^{-2} + k_2\,\tilde{r}'^{-2/3}(1-4\tilde{r}\tilde{r}'^{-2})^3) \tag{3.73}$$

where $k_1$ and $k_2$ are constants determined by the desired eye opening. For the eye opening of 60%, $k_1$=0.52 and $k_2$=0.45. For $\tilde{r}'=0$ (3.73) simplifies to (3.65) and for $\tilde{r}=(\tilde{r}'/2)^2$ it simplifies to (3.61) and also matches with the exact solution of (3.67) when $0<\tilde{r}<(\tilde{r}'/2)^2$. Comparing (3.73) with (3.61) shows that for $\tilde{r}<(\tilde{r}'/2)^2$, (3.61), the low-loss approximation for the bit rate, underestimates the bit rate of the transmission line, and for $\tilde{r}>(\tilde{r}'/2)^2$, (3.61) overestimates the bit-rate of the transmission line. The condition $0<\tilde{r}<(\tilde{r}'/2)^2$ for the structure shown in Figure 3.12, occurs when $a_R < 2Z_0\sqrt{\varepsilon/\mu}$ and $\tilde{r}>(\tilde{r}'/2)^2$ occurs when $a_R > 2Z_0\sqrt{\varepsilon/\mu}$.

Previous models suggest that the bit-rate limit is proportional to the $A/x^2$ [16, 17] or $P/x^2$ [18], where $A$ and $P$ are the cross-sectional area and perimeter of the wire. According to those models, the bit-rate limit is scale-invariant. However, in the complete

solution given by (3.65), only the $\tilde{r}'^2 T$ term is scale-invariant and the other term $\tilde{r}'^{2/3} T$ changes if all dimensions scale proportionally. Therefore, the bit-rate limit of wires is scale-invariant only if the first term in (3.65) is dominant. In other words, only when the bit-rate limit is much larger than the reciprocal $T_F$ does scaling all dimensions not affect the bit-rate limit.

### 3.5.5 Impact of ASE on the Bit-Rate Limit of Electrical Interconnects

Generality is one of the advantages of this new model. Equation (3.73) can be used to find the bit rate of any transmission line whose resistive component changes by frequency as described by (3.53) including the anomalous skin effect (ASE). As we discussed in Section 3.4, the resistivity per unit length of a wire with the structure shown in Figure 3.7 can be written as (3.52) with parameters given in Table 3.1 (on page 51). In this table, $x_{20G}$ is the length at which the bit-rate limit of a wire is 20 Gb/sec. Also, $x_{20G\text{-}ASE}$ and $x_{20G\text{-}SE}$, respectively, are the exact solution of $x_{20G}$ considering ASE and the skin effect, while $x'_{20G}$ is the conventional low-loss approximation of $x_{20G}$ based on (3.61). As was studied in the previous section, $x'_{20G}$ overestimates $x_{20G}$, because $a_R$=1. The difference between $x_{20G\text{-}ASE}$ and $x_{20G\text{-}SE}$, at small wire sizes, is mainly due to the increase in the dc resistance resulting from the surface scattering.

### 3.5.6 Bit-Rate Limit of Metal Clad Polymer Pins

According to the International Technology Roadmap for Semiconductors (ITRS), high-performance chips at the 18 nm generation will dissipate 200 W, consume greater than 300 A of supply current, and require chip-to-substrate clock frequencies greater than 80 GHz [7]. Chip input/output (I/O) interconnects provide the mechanical interconnection between the die and the package substrate or between dice that are in a 3D stack. At such

high power dissipations, it is critical to have high-density electrical I/O interconnects to minimize on-chip IR drop across the global power distribution network. With respect to signaling, an inadequate number of electrical signal I/Os limit the aggregate off-chip bandwidth, while losses due to the organic substrate and impedance mismatches and crosstalk are exacerbated as off-chip bandwidth per channel increases and the signal noise budget decreases.

A promising solution to the above needs is the use of wafer-level batch processing to fabricate mechanically compliant and high-density electrical and optical I/O interconnects. The fabrication, assembly, and basic testing of optical and dual-mode electrical-optical polymer pin I/O interconnects is described in [53]. Low loss microscopic polymer pins that are compatible with electrical solder bump fabrication provide optical I/O interconnects. Moreover, dual-mode I/Os are possible by using metal-coated polymer pins that provide both an electrical and optical path between the chip and the substrate. There is a trade-off between the electrical and mechanical compliance of the dual-mode pins. To understand this trade-off, the minimum thickness of metal needed to enable each I/O to operate at its electrical bit-rate limit needs to be studied.

In this section, we derive the minimum thickness of metal needed on a polymer pin to operate at the maximum bit-rate limit. This is important to maximize the mechanical compliance of the I/Os. SEM images of polymer pins are shown in Figure 3.13. This figure shows the pitch between two adjacent shells, $D$, the length of the polymer pins, $x$, the outer radius of pins, $r_2$, and the thickness of metal coating, $t$. The capacitance and inductance per unit length of the I/Os can be expressed as

$$\begin{aligned} c &= \pi\varepsilon \big/ \cosh^{-1}(D/2r_2) \\ l &= \mu \cosh^{-1}(D/2r_2) \big/ \pi \end{aligned} \tag{3.74}$$

where $\mu$ is the magnetic permeability and $\varepsilon$ is the electric permittivity. Therefore, the characteristic impedance of the interconnect is

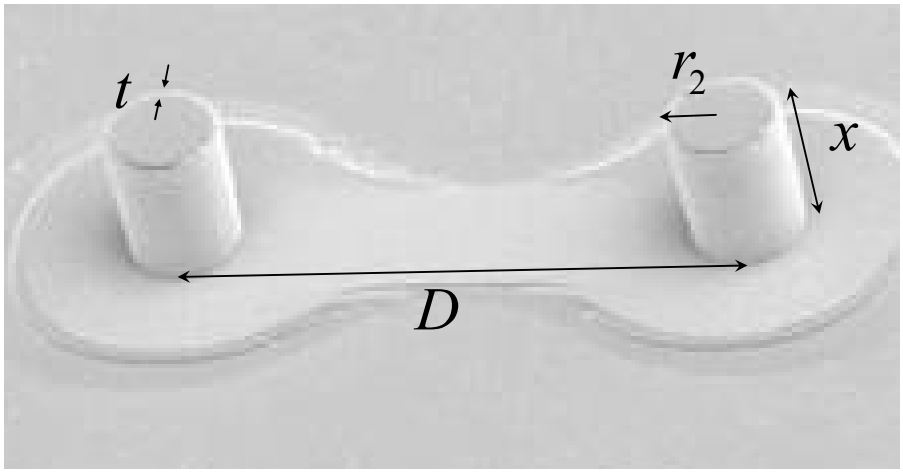



Figure 3.13: SEM images of polymer pins. The pitch between two adjacent shells, $D$, the length of the polymer pins, $x$, the outer radius of pins, $r_2$, and the thickness of metal coating, t, are shown

$$Z_0 = \frac{1}{\pi}\sqrt{\frac{\mu}{\varepsilon}}\cosh^{-1}(D/2r_2) \tag{3.75}$$

The resistance per unit length of the wire can be written as (3.35). Assuming that the bit-rate limit of the pins is fast enough, the dc term can be neglected when compared to the frequency dependent part, or $\mathbf{r}(s) = r + r'\sqrt{s} \approx r'\sqrt{s}$. For the structure shown in Figure 3.13,

$$\left.\begin{aligned} r'\sqrt{f} &= \frac{2}{\sigma}\frac{1}{2\pi r_2\delta} \\ \delta &= 1/\sqrt{\pi\mu\sigma f} \end{aligned}\right\} \Rightarrow r' = \frac{\sqrt{\pi\mu\sigma}}{\pi r_2\sigma} \tag{3.76}$$

where $\sigma$ is the electrical conductivity and $\delta$ is the skin depth. Using (3.55) and (3.61) the normalized bit-rate limit for a 60% eye opening, assuming low-loss approximations, can be written as

$$\tilde{B}_{60\%} = 0.52\tilde{r}'^{-2} = \frac{0.52}{\pi}\sqrt{\frac{\mu}{\pi}}\ \frac{\sigma r_2^{\,2}}{x}(\cosh^{-1}(D/2r_2))^2 \tag{3.77}$$

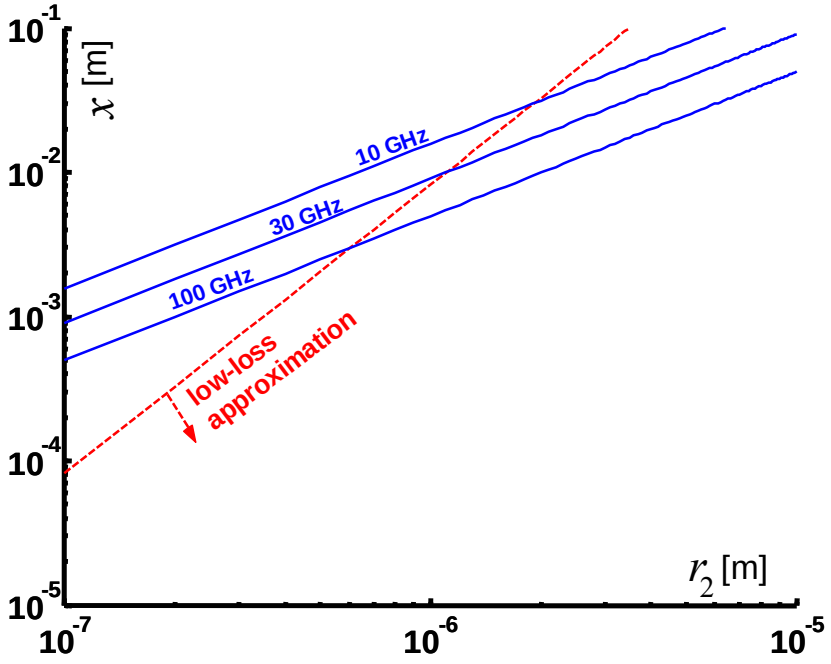



Figure 3.14: Illustrates the region of validity for the low-loss approximation and the bit-rate limit line at 10, 30 and 100 GHz.

The low-loss approximation is valid when the normalized bit-rate limit is larger than one ($\widetilde{B} > 1$). Multiplying the normalized bit rate limit by the time-of-flight, and assuming that $D/2r_2$=2.5, the bit-rate limit can be reduced to

$$B_{60\%} = \frac{\tilde{B}_{60\%}}{T_F} = \frac{0.52}{\pi} \frac{(\cosh^{-1}(D/2r_2))^2}{\rho\varepsilon} \cdot \left(\frac{r_2}{x}\right)^2 = 2.45\times10^{18}\left(\frac{r_2}{x}\right)^2 \qquad (3.78)$$

Therefore, in order for the I/Os to attain a bit-rate limit in the GHz region, the maximum aspect ratio of the pin (length to radius) should be on the order of 10,000, which is approximately three orders of magnitude larger than what is possible to fabricate and needed for our application. As such, the I/O interconnects under consideration do not impose any limitations on the signal bandwidth. Figure 3.14 illustrates the region of validity for the low-loss approximation and the bit-rate limit line at 10, 30, and 100 GHz.

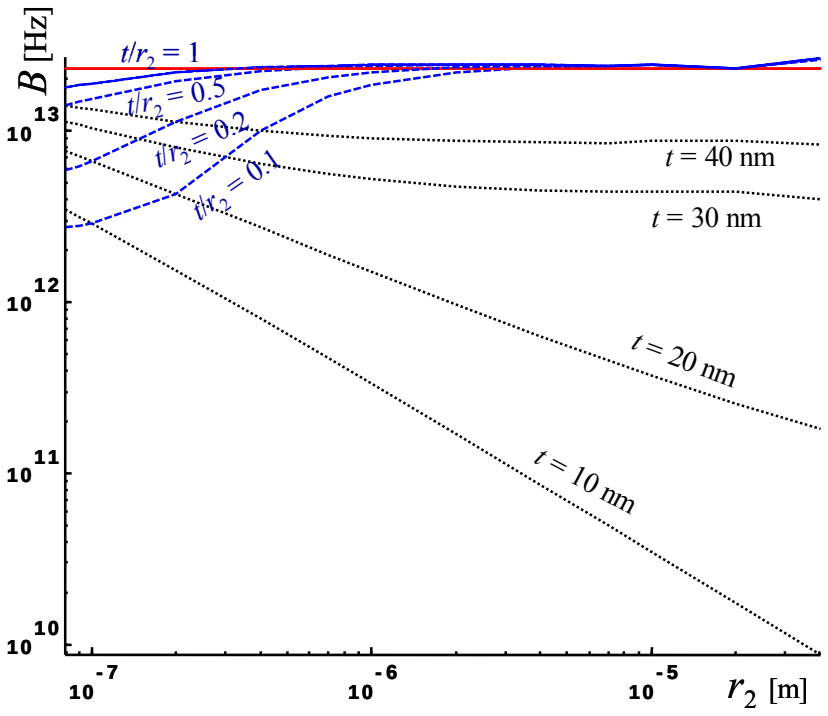



Figure 3.15: Bit-rate limit vs. wire dimension for different metal thicknesses, compared with the low-loss approximation (solid line).

Figure 3.15 is a plot of the bit-rate limit versus wire radius for different metal thicknesses ($t$). We chose two different scenarios. In the first, the metal thickness remains constant for different wire radii, and in the second, it remains a fraction of the wire's radius ($t/r_2$=1, 0.5, 0.2, 0.1). The solid line is the low-loss approximation, as in (3.78), where $x/r_2$ is constant and therefore the bit rate is independent of metal thickness. The dashed line set describes the case where $t/r_2$ remains constant. Finally, the dotted line set is for constant metal thicknesses (results of solving (3.57)). In the above analysis, it was assumed that $x/r_2$=200 and $D/2r_2$=1.5. Clearly, for aspect ratios less than 200 ($x/r_2 < 200$), the low-loss approximation is valid for I/Os with metal thicknesses as low as 50 nm. The bit-rate limit of the dual-mode polymer pins mostly depends on the ratio of

$x/r_2$ and is less dependent on the ratio of $D/r_2$. From the above analysis, a 50-nm thick metal film is the minimum thickness needed to ensure operation at the bit-rate limit of the interconnects. However, to ensure good assembly and joining to solder, where the thickness of the inter-metallic compound (IMC) between solder and copper can be 200-500 nm thick, the thickness of the conductor must be at least one order of magnitude larger. Thus, it is clear that in high-frequency signaling, the dual-mode pins will not impose a bottleneck, as they will operate up to their bit-rate limit.

### 3.5.7 Conclusions

A distributed model based on the quasi-TEM approximation is derived using several assumptions about the propagation of the fields. Analysis shows that these assumptions will be valid if the longitudinal electric and magnetic fields are much smaller than their transverse components. For on-chip GSI applications, the quasi-TEM model holds for signal frequencies up to 2.4 THz. Therefore, an interconnect in its most general form can be represented by a distributed *zy*. A phasor analysis technique is developed to model a transmission line with general *z* and *y* and to study the impact of size effects and ASE on the delay and bit-rate limit of copper interconnects.

The delay analysis shows that for different wire sizes, three regions can be defined. In region 1, the delay is much larger than time-of-flight; therefore, only the low-frequency component of the resistivity is important. This means that the main reason for the increase in the delay is surface scattering. In region 3, the delay is time-of-flight, even by increasing the resistance by ASE. Region 2 is the region between regions 1 and 3 in which the results of the ASE are slightly different from the results of considering the surface scattering and skin effect separately. Generally, for delay calculation of on-chip interconnects ASE can be neglected.

Compact models are presented for the bit-rate limit of a transmission line wherein resistance is written as the summation of dc resistance and frequency-dependent resistance. Skin effect, surface scattering, and dc resistance are hence considered simultaneously. Using a normalization method, it is shown that all equations can be written in terms of dc and ac resistances. The bit-rate limit is then identified neglecting dc resistance without making the low-loss approximation that the previous models have used, showing that the previous models underestimate the bit-rate limit by 80% [19] (when bit-rate limit is around the reciprocal time-of-flight.) In contrast to the previous models, it is shown that the bit-rate limit of a strip line remains constant with scaling of its physical dimensions only if it has an aspect ratio of 0.3. Neglecting the secondary effects such as surface scattering, down-scaling the dimensions of a strip line increases its bit-rate limit if its aspect ratio is smaller than 0.3, and increases its bit-rate limit if its aspect ratio is larger than 0.3. It is shown that for on-chip interconnects with bit-rate limits over 20 GHz, the ASE should be considered.

The dual-mode electrical-optical polymer pin is one of the promising candidates for chip input/output interconnects. The formula for the bit-rate limit of interconnects is used to study the trade-off between the electrical and mechanical compliance of the dual-mode metal clad polymer pins. It is shown that the minimum metal thickness needed to ensure operation at the bit-rate limit of interconnects is 50 nm, which is about one order of magnitude less than what is needed to ensure good assembly and joining to the solder. Thus, in high-frequency signaling, the dual-mode pins will not impose a bottleneck, as they will operate up to their bit-rate limit.

# CHAPTER 4

# Impact of Size Effects on the Design and Performance of Metal Interconnect Networks

## 4.1 Introduction

Some of the most significant limitations for gigascale integration (GSI) are posed by interconnect networks [3]. The performance of interconnects, unlike transistors, does not improve by miniaturization. Reverse scaling in the multi-level interconnect network is used to ease this problem and to satisfy the set of performance requirements [54]. About a decade ago, copper replaced aluminum as an advanced metallization solution [5]. Besides higher conductivity, copper interconnects offer good mechanical strength, a higher melting point, and better electromigration performance [5, 55]. The mean free path of the electrons for copper at room temperature is about 40 nm. As wire dimensions and grain size become comparable to the mean free path of the electrons, resistivity increases because of surface scattering, grain boundary scattering, and surface roughness. The 2006 ITRS projects that for the year 2020, interconnects will be as narrow as 14 nm and the resistivity of minimum-size interconnects will be more than four times larger than the bulk resistivity [56].

Size effects have been the subject of many studies in the past [28, 38, 57-59], and currently many research groups are investigating techniques that can potentially mitigate size effects to improve copper conductivity for future technology generations. It has been shown that size effects exacerbate the thermal problems that multi-level interconnect networks face because of scaling [60]. The impact of wire resistivity on dynamic system

level performance has been studied using a ring oscillator [61]. These results show that wire resistance determines the circuit performance only when it contributes 10% or more to the total delay of one stage of the ring, and the wire capacitance plays a greater role in circuit performance.

However, a multi-level interconnect network consists of interconnects with a wide range of lengths that are routed in metal levels with different pitches and thicknesses. The impact of size effects on the design of multi-level interconnect networks has not been studied, and it is not clear how much overall chip performance will be degraded as a result of size effects. Such quantitative analyses can be very helpful for any kind of cost/performance analysis for various size-effect mitigation techniques.

Historically, utilizing more metal levels has been a solution to interconnect problems [62]. In this chapter, after a brief review of two different design paradigms for high-performance and low-cost designs, a multi-level interconnect network is designed considering surface and grain boundary scatterings. The results are then compared with a hypothetical case in which size effects do not exist to quantify the impact of size effects on the number of metal levels and chip performance.

## 4.2 Mutli-level Interconnect Network Design

To provide optimal design for a multi-level interconnect network, the wiring distribution model described in [63] is used. The continuous interconnect density function, $i(l)$, is defined such that $i(l)dl$ is the number of point-to-point interconnects whose length is $l \pm dl/2$, where $l$ is the interconnect length normalized to the average gate pitch. Using the recursive application of Rent's rule, $i(l)$ is given as [63]

$$\begin{aligned}
&\text{Region I}: \quad 1 \le l \le \sqrt{N_g} \\
&i(l) = \frac{\alpha k_R}{2}\Gamma\left(\frac{l^3}{3} - 2\sqrt{N_g}l^2 + 2N_g l\right) l^{2p_R - 4} \\
&\text{Region II}: \quad \sqrt{N_g} \le l \le 2\sqrt{N_g} \\
&i(l) = \frac{\alpha k_R}{6}\Gamma\left(2\sqrt{N_g} - l\right)^3 l^{2p_R - 4}
\end{aligned}, \tag{4.1}$$

where $N_g$ is the number of logic gates, $p_R$ is the Rent's exponent, $k_R$ is the Rent's coefficient, $\alpha$ is the fraction of sink terminals in the macrocell, and $\Gamma$ is the normalizing factor. A gate pitch is defined as the average distance separating two gates and is equal to $\sqrt{A_m / N_g}$ , where $A_m$ is the macro-cell area.

The impact of size effects on the design and performance of multi-level interconnect networks in two different design paradigms is studied. For high-performance chips, the wiring pitch in each metal pair can be optimized such that the longest interconnect in each pair meets a delay constraint (n-tier methodology), while for low-cost chips very few wiring pitches are normally used to avoid multiple recipes that are expensive to develop and maintain (two-tier methodology) [64].

### 4.2.1 N-Tier Multilevel Architecture Design

The n-tier methodology optimally designs the wire pitch of each orthogonal pair of metal levels such that the longest interconnect in each level meets the given time constraints [54]. Using the interconnect density function (4.1), the relation between the available wiring area (supply) and required wiring area (demand) for each layer can be written as

$$e_w A_m = \chi p_L \sqrt{A_m / N_g} \int_{L_{\min}}^{L_{\max}} l i(l) dl \tag{4.2}$$

where $e_w$ is the wiring efficiency factor, $\chi$ is the point-to-point conversion factor, and $p_L$, $L_{\min}$ , and $L_{\max}$ are the wire pitch, the shortest, and the longest interconnect lengths on the levels, respectively. In this work, following [4], it is assumed that $e_w$ =40% and $\chi$=0.67.

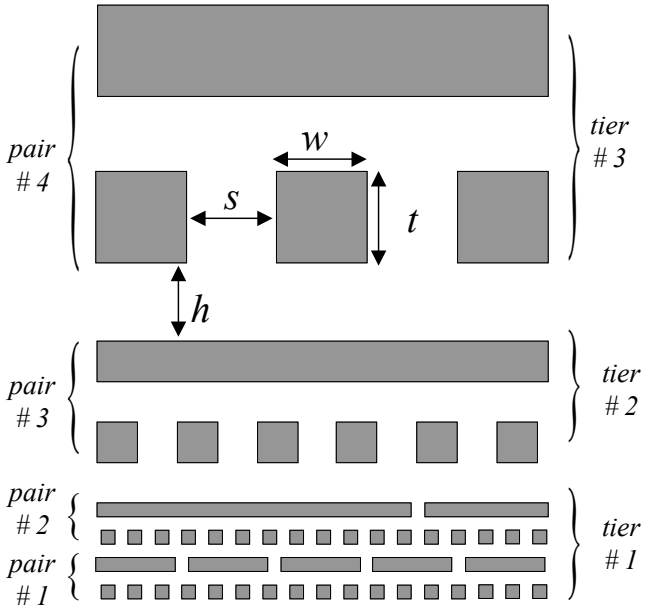



Figure 4.1: Cross-sectional view of n-tier design methodology with eight levels, four pairs and three tiers.

Two adjacent levels with orthogonal metal levels are called a *pair*, and a *tier* is a collection of pairs with the same pitch. It is assumed that all the cross-sectional dimensions of wires are equal: $w=t=s=h=p/2$, where $w$, $t$, $s$, and $h$ are the metal width, wire thickness, spacing between interconnects, and height of the inter-level dielectrics, respectively, as shown in Figure 4.1.

The *rc* time delays of interconnects, without and with repeater insertion, are given by (4.3.a) and (4.3.b), respectively:

$$\tau_{rc} = 1.1 rc l^2 \tag{4.3.a}$$

$$\tau_{rc} = \left(1.4 + 0.53\zeta + 0.53/\zeta\right)\sqrt{R_0 C_0 rc}\, l \tag{4.3.b}$$

where $r$ and $c$ are resistance and capacitance per unit length of the interconnect, $R_0$ and $C_0$ are output resistance and input capacitance of a minimum-size inverter, and $\zeta$ is the sub-

optimal factor of repeater insertion, which is assumed to be 50% throughout this work [54].

Starting from the first pair of wiring levels, (4.2) and (4.3) are solved simultaneously to determine the minimum wiring pitch that can satisfy the time constraint for the longest interconnect in this pair. If this wiring pitch is smaller than the minimum pitch projected by the ITRS, the ITRS minimum pitch is used, and using (4.2), the length of the longest interconnect is calculated. The maximum length for Pair 1 becomes the minimum length for Pair 2, and the same steps are repeated for all pairs until the longest interconnect in the wiring distribution is routed. This ensures that the time delay constraints are exactly satisfied and also that the wiring density is maximized. The timing constraint here is determined as $\tau_{rc} = \beta / f_c$ , while $\beta$ is chosen to be 0.25 for local interconnects and 0.9 for longer interconnects [54].

### 4.2.2 Two-Tier Multilevel Architecture Design

The two-tier methodology optimally designs the partition length between the tiers and the wiring pitch in the second tier such that for a given number of metal levels, the average interconnect delay is minimized. In low-cost designs, the same BEOL process is usually used for many applications; hence there is no information about interconnects that are routed in the critical path. However, the probability of interconnects with length $l \pm dl$ being in the critical path is proportional to the number of interconnects with length $l \pm dl$. The average critical path delay can be written as

$$\left\langle \tau_{critical} \right\rangle = D_L \left\langle \tau \right\rangle , \tag{4.4}$$

where $D_L$ is the logic depth of the critical path and $\langle \tau \rangle$ is the average delay of all wires. Therefore by minimizing the average interconnect delay, the average critical path delay will be minimized.

The area constraint in tier 1 and 2 will be written as

$$2N_1 e_w A_m = \chi p_1 \sqrt{A_m / N_g} \int_0^{L_{par}} li(l)dl \,, \tag{4.5.a}$$

$$2N_2 e_w A_m = \chi p_2 \sqrt{A_m / N_g} \int_{L_{par}}^{L_{\max}} li(l)dl \,, \tag{4.5.b}$$

where $N_1$ and $p_1$ are the number of metal level pairs and the wire pitch in tier 1, and $N_2$ and $p_2$ are the number of metal level pairs and the wire pitch in tier 2, respectively. The two-tier methodology finds $p_2$ and $L_{par}$ such that for a given number of metal pairs, $N$ (=$N_1$+ $N_2$), $\langle \tau \rangle$, the average delay of all wires is minimized.

$$\langle \tau \rangle = \int_0^{L_{\max}} \tau(l) i(l) dl \,, \tag{4.6}$$

where $\tau$, interconnect delay, is as given in (4.3).

## 4.3 Impact of Size Effects on Copper Resistivity

The *rc* time delay of interconnects, as given in (4.3), is a function of resistance per unit length of the interconnects. The copper resistivity for interconnects implemented in future technology generations will be larger than copper bulk resistivity, as wire cross-sectional dimensions become comparable to the mean free path of electrons. Two major scattering effects are *surface scattering* and *grain boundary scattering*. The Fuchs-Sondheimer model presented in [14] is used to model the surface scattering. In this model, an empirical parameter *p*, called the specularity parameter, is used to describe the fraction of electrons scattered specularly at the surface. The resistivity of a wire with unity aspect ratio can be written as

$$\frac{\rho}{\rho_b} = 1 + \frac{3}{4}(1-p)\frac{\lambda}{w}, \tag{4.7}$$

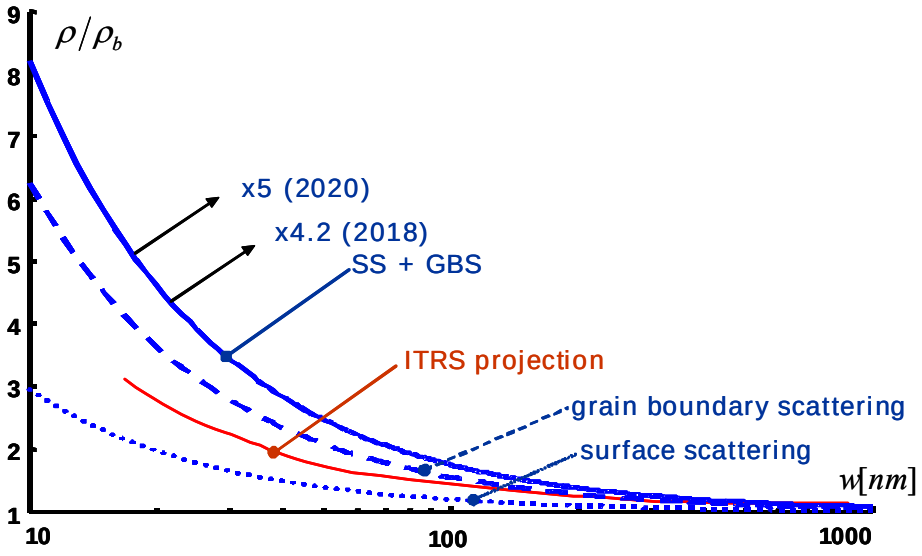



Figure 4.2: Copper resistivity (normalized to bulk resistivity) versus wire dimension considering surface scattering (dotted line), grain boundary scattering (dashed line) and both effects combined (solid line) compared with the ITRS projections

where $\rho_b$ is the bulk resistivity, $\lambda$ is the mean free path of the elections, and $w$ is the wire width.

The grain boundary scattering is modeled as the Mayas-Shatzkes method [65]. In this model, another empirical parameter *R*, called the reflectivity coefficient, describes the probability of electrons getting reflected at the grain boundaries. Based on this model, resistivity can be written as

$$\frac{\rho}{\rho_b} = \frac{1}{3}\left(\tfrac{1}{3} + \tfrac{\alpha}{2} + \alpha^2 - \alpha^3 \ln\left(1 + \tfrac{1}{\alpha}\right)\right)^{-1}, \quad as \quad \alpha = \frac{R}{1-R}\frac{\lambda}{D} \tag{4.8}$$

where *D* is the grain diameter. In this work, the size of the grains is assumed to be equal to the wire width (*D*=*w*), and empirical parameters *p* and *R* are both chosen to be 0.5 [59]. The resistivity of copper interconnects versus wire dimension, based on these assump-

tions, is plotted in Figure 4.2 and compared with the ITRS projections for the resistivity of minimum-size wires. For this study the chosen values are pessimistic enough.

Table 4.1: Data used from ITRS for case study

| Year | 2003 | 2018 |
|---|---|---|
| ½ wiring pitch (*nm*) | 120 | 21 |
| clock frequency $f_c$ (*Ghz*) | 2.9 | 53.2 |
| chip area $A_m$ ($cm^2$) | 5.79 | 5.79 |
| number of transistors $N$ (*millions*) | 810 | 25915 |

## 4.4 Metal Interconnect Networks for High-Performance Chips

The n-tier methodology explained in Section 4.2.1 is used in this section to design the interconnect network of two high-performance chips implemented in 2003 and 2018, which correspond to the ITRS 100 nm and 18 nm nodes (Table 4.1). For the 18 nm node, it has been assumed that the chip consists of 36 macrocells such that all intra-macrocell signals can be transferred in one clock cycle [66]. Assuming that the inter-macrocell communication is through a regular network, the n-tier methodology is applied to intra-macrocell interconnects only.

### 4.4.1 Impact of Size Effects on Design of N-Tier Architecture

The n-tier architecture is designed once, assuming there are no size effects (CASE I), and then the n-tier architecture is redesigned considering size effects (CASE II). The results of these two designs for the 100 nm and 18 nm technology nodes are shown and com-

pared in Tables 4.2 and 4.3, respectively. Interestingly, in 2018, while the resistivity for minimum-sized wires in *CASE II* is 4.4 times higher than the bulk resistivity used in

Table 4.2: Results of n-tier design methodology neglecting (CASE I) and considering (CASE II) size effects for year 2003 (100 nm node). $L_{max}$ is normalized to the gate pitch (840 nm)

| | *CASE I* | | *CASE II* | | *Difference* | |
|---|---|---|---|---|---|---|
| *# of pairs (Np)* | 4.0563 | | 4.0812 | | 0.6 % | |
| | *Pitch* (*nm*) | $L_{max}$ | *Pitch* (*nm*) | $L_{max}$ | *Pitch* (*nm*) | $L_{max}$ |
| *pair #1* | 240 | 6 | 240 | 6 | 0 % | 0 % |
| *pair #2* | 240 | 48 | 240 | 48 | 0 % | 0 % |
| *pair #3* | 240 | 1495 | 240 | 1495 | 0 % | 0 % |
| *pair #4* | 717 | 23888 | 724 | 21958 | 0.9% | -8% |
| *pair #5* | 1709 | 56920 | 1780 | 56920 | 4.1% | 0 % |

CASE I (Figure 4.2), the designs for these two cases are not drastically different. This is mainly because most interconnects that are largely affected by size effects are so short that their resistances do not play an important role. This can be explained in more detail through Table 4.3.

All interconnects routed in the first two pairs are so short that the minimum wire width meets the time constraint even when size effects are considered. For Pair 3, the wire width in CASE II has to increase to compensate the larger resistivity compared to CASE I. This increase, however, is quite moderate because a slight increase (9.5%) in wire pitch reduces the maximum length in Pair 3 by a larger percent (29%), which again compensates the increase in resistivity. The minimum and maximum lengths for Pair 4, therefore, are considerably smaller in CASE II than the ones in CASE I (29% and 34%, respectively), and accordingly, the required pitch is smaller. This trend continues through all higher pairs. Note that in Pair 4 and above, the wire dimensions are several times

larger than the electron mean free path, and size effects are hence less severe (e.g., $\rho/\rho_b$=1.97 in Pair 4 and $\rho/\rho_b$=1.37 in Pair 5).

Table 4.3: Results of n-tier design methodology neglecting (CASE I) and considering (CASE II) size effects for year 2018 (18 nm node). $L_{max}$ is normalized to the gate pitch (350 nm)

| | *CASE I* | | *CASE II* | | *Difference* | |
|---|---|---|---|---|---|---|
| *# of pairs (Np)* | 6.004 | | 6.4044 | | 6.7 % | |
| | *Pitch* (*nm*) | $L_{max}$ | *Pitch* (*nm*) | $L_{max}$ | *Pitch* (*nm*) | $L_{max}$ |
| *pair #1* | 42 | 6 | 42 | 6 | 0 % | 0 % |
| *pair #2* | 42 | 44 | 42 | 44 | 0 % | 0 % |
| *pair #3* | 42 | 1077 | 46 | 760 | +9.5 % | -29 % |
| *pair #4* | 152 | 5125 | 143 | 3389 | - 5.9 % | -34 % |
| *pair #5* | 406 | 13683 | 326 | 9120 | - 20 % | -33 % |
| *pair #6* | 1199 | 40388 | 698 | 21353 | - 41 % | -47 % |
| *pair #7* | 1593 | 53660 | 1664 | 53660 | +4.4 % | 0 % |

In summary, the short interconnects in the first pair are minimum-width limited, and size effects do not change their design. The increase in resistivity as a result of size effects requires a slightly larger wire pitch in Pair 3, which reduces the number of interconnects that can be routed in this pair. The burden of this reduction in wiring density of Pair 3 becomes shared among all higher pairs in which size effects are less severe. The increase in the number of metal levels resulting from size effects for the case study at the 18 nm node will be 0.4 pairs (6.7%).

Throughout this work Rent's exponent, $p_R$, and Rent's coefficient, $k_R$, are chosen as 0.55 and 4, respectively [67]. Although the wiring distribution and therefore the required number of metal levels depends on Rent's parameters, the difference between the total number of metal levels in CASE I and CASE II does not depend on Rent's parameters. The increase in the number of metal levels in CASE II (relative to CASE I) for different

values of Rent's parameters is plotted in Figure 4.3. For the Rent's exponents from 0.5 to 0.6 and the Rent's coefficients from 3.4 to 4.8, the relative increase in the total number of metal levels, due to size effects, remains between 2% and 17%.

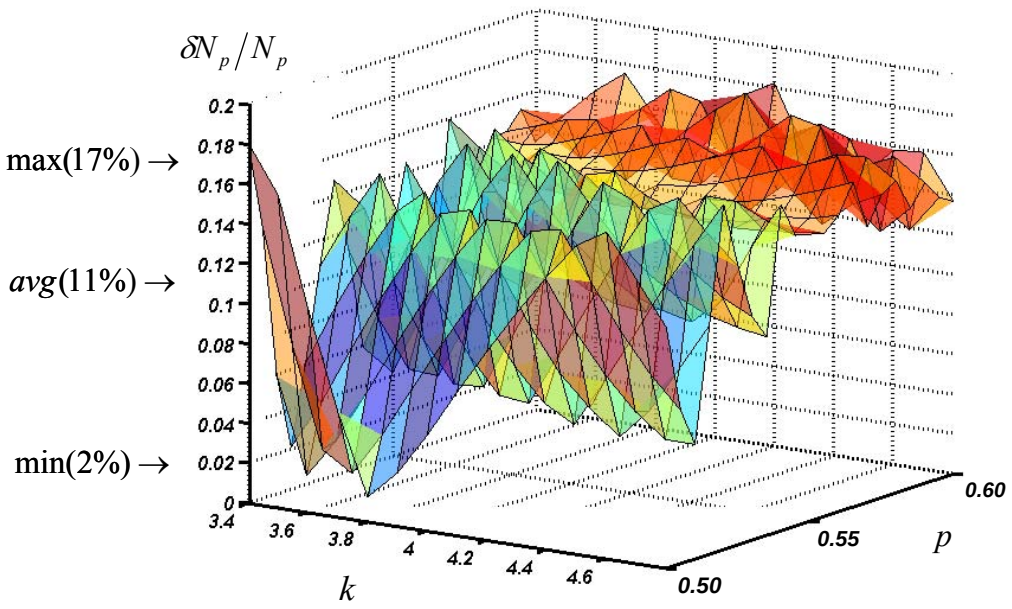



Figure 4.3: Relative increase in the total number metal levels needed to address size effects versus different values of Rent's exponent and Rent's coefficient.

### 4.4.2 Impact of Size Effects on Performance of N-Tier Architecture

Even though maximum interconnect latencies in CASE I are the same as the ones in CASE II, the latency distributions are different in the two cases. This means that a critical path that consists of interconnects with many different lengths may have a larger latency once size effects are present (CASE II). To quantify the impact of size effects on overall chip performance, one should look at the latency distribution in two cases. Starting from wiring length distribution, $i(l)$, and by using the n-tier methodology, a

wiring *rc* delay distribution, $i(\tau)$, is plotted in Figure 4.4. It shows that interconnects that are most affected by size effects are so short that their *rc* time delay is negligible compared to the delays of longer interconnects. In 2018, the minimum intrinsic delay of a gate will be $\tau_0$=0.21*ps*, roughly one percent of the clock cycle. Based on this, we categorize interconnects into two groups. Group A includes interconnects whose *rc* delay is less than 0.01/*fc*, and group B are those whose *rc* delay is larger than 0.01/*fc*. Intuitively, the time delay of interconnects in group A is so small that any increase in their latencies has a negligible impact on the overall chip performance.

The average latencies of total interconnects and interconnects in groups A and B are shown in Table 4.4 for three different cases. The first row is for a hypothetical case where

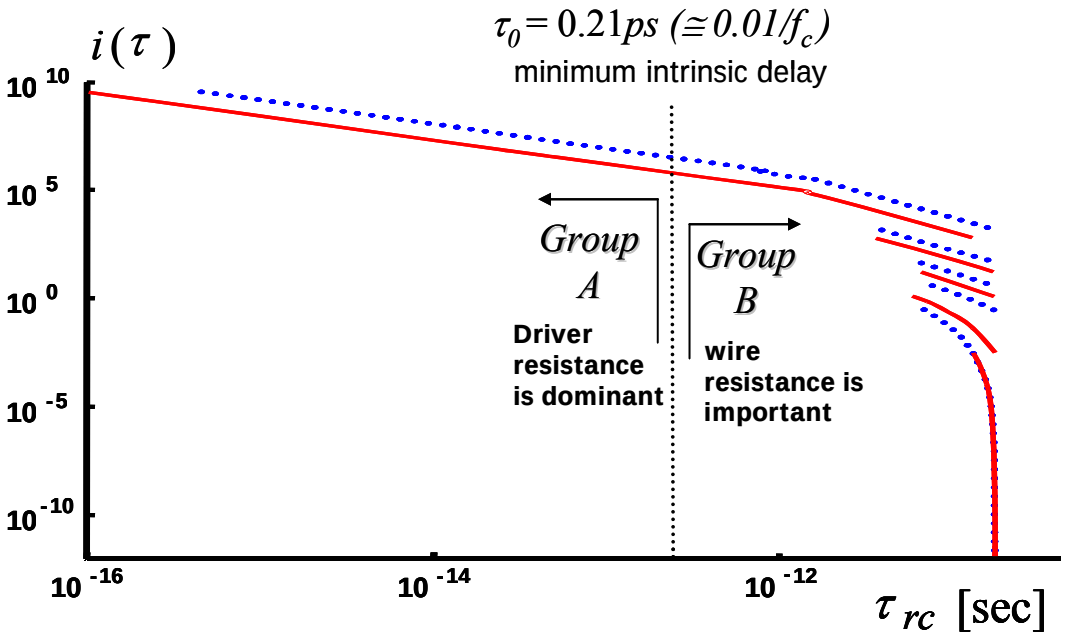



Figure 4.4: Wire *rc* delay distribution for a case study of year 2018 from ITRS using the n-tier methodology; the solid line is CASE I, neglecting size effects, and the dotted line is CASE II, considering size effects. *rc* delay of group A wires (routed in the first tier) is smaller than the minimum intrinsic delay.

Table 4.4: $\langle \tau_{rc} \rangle$ average *rc* time delay of all interconnects, $\langle \tau_A \rangle$, average *rc* delay of wires with *rc* delay less than $0.01/fc$, and $\langle \tau_B \rangle$, those with *rc* delay of more than $0.01/fc$.

| | $\langle \tau_{rc} \rangle$ (*fsec*) | $\delta\tau/\tau$ | $\langle \tau_A \rangle$ (*fsec*) | $\delta\tau/\tau$ | $\langle \tau_B \rangle$ (*psec*) | $\delta\tau/\tau$ |
|---|---|---|---|---|---|---|
| *CASE I* | 9.91 | - | 1.69 | - | 1.377 | - |
| *CASE II* | 23.3 | 135% | 3.98 | 136% | 1.308 | -5 % |
| *CASE III* | 26.5 | 167% | 3.98 | 136% | 1.521 | 10 % |

there are no size effects (CASE I). The second row corresponds to the design that is based on size effects. The third row is for a design in which size effects are present but are ignored in the design process (CASE III). So, wire dimensions for the first and third rows are the same. Table 4.4 shows that size effects increase the average rc time delay of all interconnects, $\langle \tau_{rc} \rangle$, by 135% if they are considered in the design process and by 165% if they are ignored. However, the changes are mainly due to interconnects in group A that have negligible latencies. The average latency of the interconnects in Group B, however, will increase by only 10% even if size effects are ignored in the design process, and it will slightly decrease (5%) if the interconnect network is designed taking into account size effects. This decrease in the average latency is because many interconnects that would have been routed in Pair 3 will be routed in Pair 4, with a much larger pitch.

Overall, for high-performance applications, the results of this section indicate that the impact of size effects on chip performance will be quite limited, especially if size effects are considered during the design process. Figure 4.5 summarizes the impact of size effects on the n-tier architecture design. The relative increase in the average inter-connect delay, $\langle \tau \rangle$, and relative increase in the number of metal levels using n-tier methodology, N, is plotted for different technology nodes.

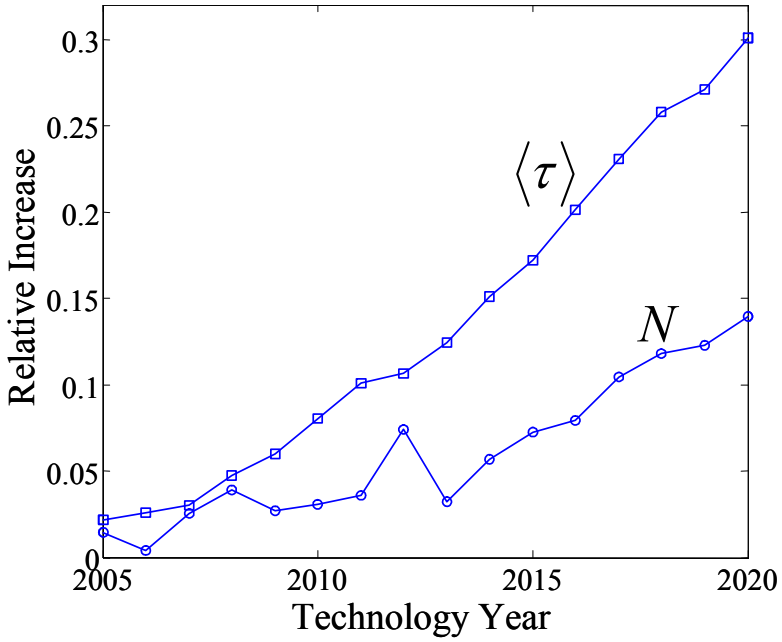



Figure 4.5: Relative increase in the average wire delay due to size effects and number of metal levels needed to compensate them versus technology year.

## 4.5 Metal Interconnect Networks for Low-Cost Chips

### 4.5.1 Impact of Size Effects on Design of Two-Tier Architecture

The two-tier methodology presented in Section 4.2.2 is used in this section to design the interconnect network for a low-cost chip implemented in 2020. At this technology node, the resistivity of minimum-size wires will be five times larger than the bulk resistivity (Figure 4.2). The two-tier architecture is designed in two different cases: CASE I, assuming there is no size effects and CASE II, considering size effects. The results of these two designs are shown in Table 4.5. The optimal design is when the first tier consists of just one metal pair routed in the minimum pitch, while the wiring pitch in the second tier is only determined by the total number of metal pairs.

Table 4.5: Results of the low-cost design methodology neglecting (*CASE I*) and considering (*CASE II*) size effects for the year 2020 (14*nm* node). $L_{par}$ is normalized to the gate pitch (230*nm*)

| | *CASE I* | | *CASE II* | |
|---|---|---|---|---|
| $N_{pairs}$ | $P_2$ (*nm*) | $L_{par}$ | $P_2$ (*nm*) | $L_{par}$ |
| *3* | 35 | 12 | 35 | 12 |
| *4* | 50 | 12 | 50 | 12 |
| *5* | 70 | 12 | 70 | 12 |
| *6* | 85 | 12 | 85 | 12 |
| *7* | 105 | 12 | 105 | 12 |
| *8* | 120 | 12 | 120 | 12 |
| *9* | 140 | 12 | 140 | 12 |

### 4.5.2 Impact of Size Effects on Performance of Two-Tier Architecture

Although the optimal design is not affected by size effects, the performance of an inter-connect network will be affected. Table 4.5 indicates that wire pitch in the second tier decreases as the total number of metal levels decreases. Hence, the low-costs chips with a smaller number of metal levels suffer more from size effects. Figure 4.6 shows the relative increase in the maximum delay, $\tau_{\max}$, and the average delay, $\langle \tau \rangle$, versus number of metal levels. Because of size effects, the maximum delay and the average delay of the wires could increase up to 100% and 40%, respectively.

In the previous section, it was shown that size effects could be addressed by utilizing more metal levels for an n-tier architecture. To use the same techniques, in Figure 4.7, the average delays of interconnects are plotted versus the number of metal pairs in CASE I and CASE II. The average delay in CASE II will be equal to the average delay in CASE I if the number of metal pairs is increased from 3 to 4.75 or from 5 to 8. Unlike the high-

performance applications in which a 13% increase in the number of metal levels was enough to compensate the impact of size effects on the performance of an interconnect network at the end of the ITRS roadmap (Figure 4.5), for low-cost applications, a 60% increase in the number of metal levels is needed to address size effects.

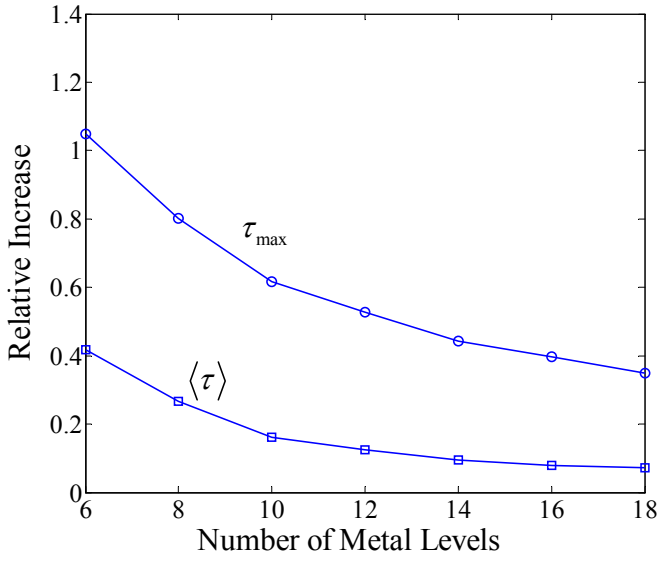



Figure 4.6: Relative increase in the average delay, $\langle\tau\rangle$, and the maximum delay, $\tau_{max}$, for a 2020 low-cost chip implemented in two-tiers versus number of metal levels.

## 4.6 Conclusions

It is shown that for high-performance applications, modest changes in the design of a multi-level interconnect network can address the increases in copper resistivity due to surface and grain boundary scatterings such that the longest interconnect in each metal level meets the specified time constraint. For a case study chip implemented at the 18 nm

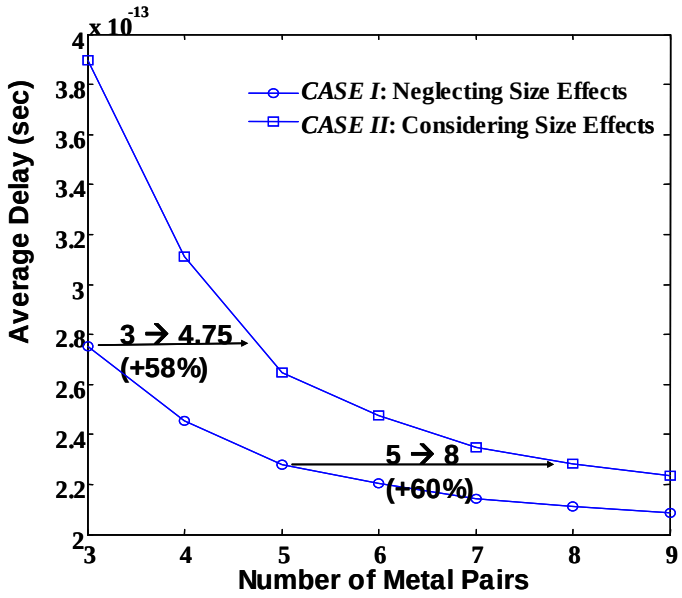



Figure 4.7: Average delay versus number of metal levels in two cases, considering and neglecting size effect for a 2020 low-cost chip implemented in two-tiers versus number of metal levels.

node, it is shown that the changes in wiring pitches and number of metal levels will be less than ±40% and 7%, respectively, while the copper resistivity for the minimum-size interconnects will increase by more than four times. The main reason for the modest changes is that interconnects that are affected most by size effects are so short that their latencies are not important. Slight changes in wiring pitches allow for transferring the critical ones that are hurt by size effects to higher metal levels with larger wiring pitches such that they meet the specified timing constraint.

A novel latency distribution is presented that shows that the average latency of all interconnects increases by 136% because of size effects. However, this increase is mainly due to the short interconnects, whose latencies are less than 1% of the clock cycle. The

average latency of interconnects with latencies larger than 1% of the clock cycle can even slightly decrease because of the larger number of metal levels utilized.

It is shown that for low-cost applications, very few wiring pitches are normally used and the number of metal levels needed to compensate the impact of size effects on the average *rc* delay of copper interconnect is drastically high. For a case study chip implemented at 14 nm with a two-tier interconnect network, it is shown that a 60% increase in the number of metal levels is needed to address a five times increase in the resistivity of minimum-size wires.

It is important to note that this analysis is based on the latency of signal interconnects only. Many other issues such as electromigration, impact of size effects on power distribution, etc., need to be investigated carefully.

# CHAPTER 5

# Design and Optimization of Local Power Distribution Network

## 5.1 Introduction

As technology advances, it becomes more and more challenging to design on-chip power distribution networks. This is because power supply voltage and hence the noise margin scale down with technology, while the number of devices in a die and the current per unit area of a die scale up [24, 25, 68]. The *IR* drop and *Ldi*/*dt* noise associated with the power distribution network are crucial to circuit timing and performance [25, 69]. Historically, the inductive behavior of power networks was dominated by off-chip parasitic inductance and the inductive behavior of on-chip power distribution networks has been neglected. However, with the introduction of advanced packaging techniques, this has changed [68]. The on-chip power line can be modeled as distributed *rlc* segments. The *l*/*r* delay characterizes the importance of inductance in power network modeling. If the *l*/*r* delay is much smaller than the *rc* delay, the inductance can be ignored in the modeling [25]. Hence, for the local power distribution network (narrow lines) *IR* drop and for the global power distribution network (wide lines) both *IR* drop and *Ldi*/*dt* should be considered in the modeling and optimization of power networks.

In [70], an algorithm based on efficient nonlinear programming techniques has been used to minimize the area of a power network for a given IR drop budget. An optimization algorithm for determining the widths of power and ground routes of a multi-pad distribution system is presented in [71]. A linear programming method is applied in [72] to find the proper sizing for power/ground networks. Although compact physical models for IR drop and simultaneous switching noise have been proposed recently in [73] and

[74] for a global power distribution network, a compact physical model and closed-form optimization technique for a local power distribution network have not yet been presented.

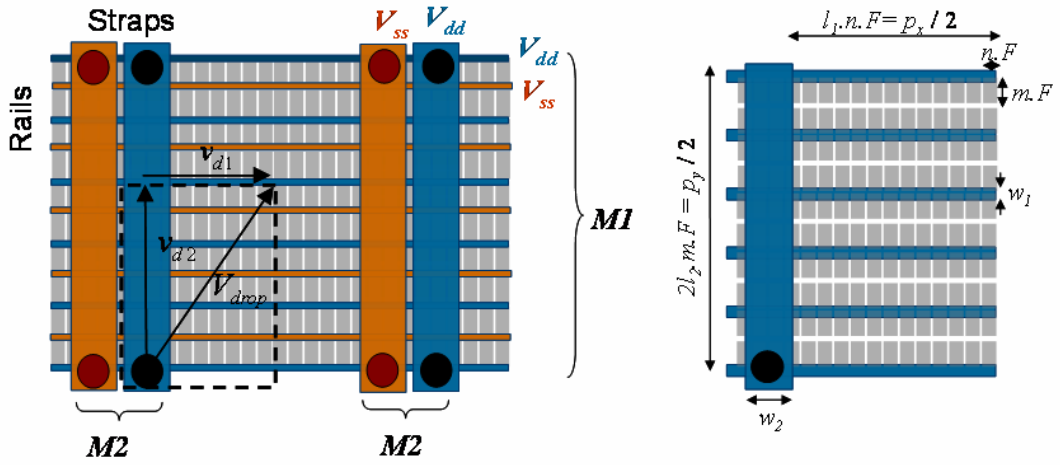
At the global level, power and ground lines are routed in the top metal levels with relatively large cross-sectional dimensions, and size effects do not affect them. At the local level, however, power and ground lines are routed in the first two metal levels to deliver power and ground to each individual gate. Because of their small dimensions, these interconnects are vulnerable to size effects. It is therefore critical to model and optimize power distribution networks at the local level and to quantify the impact of size effects on the design and performance of local power distribution networks, which are the subject of this chapter.

As the feature sizes of interconnects scale down, copper resistivity increases due to various size effects such as surface scattering and grain boundary scattering. At the end of the ITRS, local interconnects are projected to be as narrow as 14 nm [7], and the resistivity of minimum-size interconnects will be more than four times larger than the bulk resistivity of copper. Previous studies show that the impact of size effects on high-performance chips will be relatively small, especially once size effects are considered in the design process of multi-level interconnect networks [75]. However, for low-cost applications in which few wiring pitches are normally used to avoid multiple recipes, which are expensive to develop and maintain, size effects have a considerable impact on the performance of interconnect networks [76].

In this work, the local power distribution network is optimized such that the total metal area in the first two metal levels is minimized for a given *IR* drop budget. Then, the impact of size effects on the local power distribution networks is studied. The area blocked by the power vias connecting the local to the global network is studied and the optimal power via pitch, and consequently the optimal width of the local power network, is presented.

## 5.2 Assumptions and Problem Definition

A local power distribution network consists of power rails routed in metal 1 (M1), connected to orthogonal power straps routed in metal 2 (M2), as shown in Figure 5.1. The local power network is connected to the global power grid through vias. The maximum *IR* drop in the local network, $V_d$, is at the center of a region between the four nearest power vias as shown in Figure 5.1-a. This voltage drop is the summation of the *IR* drops in M1 and M2 that are called $v_{d1}$ and $v_{d2}$, respectively. Because of the symmetry, only the tile shown by the dashed square in Figure 5.1-a and "zoomed" in Figure 5.1-b needs to be modeled. The size of the tile is determined by the pitch of the global power vias in the $x$ and $y$ directions, $p_x$ and $p_y$ as shown in Figure 5.1-b. Parameters $w_1$ and $w_2$ are the widths of the power lines in M1 and M2, respectively. A typical logic gate is assumed to be an




(a) (b)

Figure 5.1: (5.1-a) Power and ground distribution networks in the first two metal levels, M1 and M2. Circles represent vias connecting local power and ground grids to the power grids in the global levels. (5.1-b) The "zoomed" tile of power rails and straps with their geometrical definitions.

$nF \times mF$ rectangle, where $F$ is the feature size. Assuming that inside each tile there are $l_1$ gates in the $x$ direction and $l_2$ power rails in the M1, the pitch of the power vias can be written as

$$\begin{aligned} p_x &= 2l_1 nF \\ p_y &= 4l_2 mF \end{aligned}, \tag{5.1}$$

Following [77], in this work it is assumed that $n$ =12, $m$ =28. Using Ohm's law and arithmetic series, the *IR* drops in M1 and M2 can be written as

$$\begin{aligned} v_{d1} &= 2I_0 \rho \frac{nF}{tw_1}(1+2+\cdots+l_1) = \frac{I_0 \rho p_x}{2tw_1}(\frac{p_x}{2nF}+1) \\ v_{d2} &= 4l_1 I_0 \rho \frac{mF}{tw_1}(1+2+\cdots+l_2) = \frac{I_0 \rho p_x p_y}{2ntw_2 F}(\frac{p_y}{4mF}+1) \end{aligned}, \tag{5.2}$$

where $I_0$ is the supply current per gate, and $t$ is the metal thickness in the first two metal levels.

The average supply current per logic gate, $I_{avg}$, is calculated by

$$I_{avg} = \frac{nF \cdot mF}{A_{chip}} \cdot \frac{P_{\max}}{V_{dd}}, \tag{5.3}$$

where $A_{chip}$ is the chip area, $P_{\max}$ is the maximum chip power, and $V_{dd}$ is the supply voltage which are all taken from the ITRS projections [7]. The direct current per logic gate, $I_0$, is assumed to be $k$ times larger than the average supply current per logic gate ($I_0 = k\, I_{avg}$ ) because not all logic blocks in a chip are active at a given time.

Using the definitions in Figure 5.1-b, the fraction of the metal area needed for the power distribution network can be written as (considering a minimum spacing of $F$ between the lines)

$$\frac{A_{power}}{A_{chip}} = \frac{w_1 + F}{2mF} + \frac{w_2 + F}{p_x}, \tag{5.4}$$

The goal is to find the optimal values for $w_1$ and $w_2$ such that for a given IR drop budget, $V_{IR}$, the total metal area for the power distribution, $A_{power}$ (5.4), is minimized.

$$\begin{cases} \min\left(A_{power}(w_1, w_2)\right) \\ V_d = v_{d1}(w_1) + v_{d2}(w_2) = V_{IR} \end{cases}, \tag{5.5}$$

For the case studies in this paper, $V_{IR}$ is chosen to be 2% of the supply voltage ($V_{IR}$ = $0.02V_{dd}$) and wire thickness, $t$, is assumed to be twice the minimum feature size (aspect ratio of 2).

## 5.3 Area Optimization in the First Two Metal Levels

### 5.3.1 Neglecting Size Effects

To solve (5.5), $A_{power}$ should be minimized with two independent variables and one constraint ($V_d$=$V_{IR}$,); therefore the evaluation of one Jacobian determinant is required.

$$J\left(\frac{A_{power}, V_d}{w_1, w_2}\right) = \begin{vmatrix} \frac{\partial A_{power}}{\partial w_1} & \frac{\partial A_{power}}{\partial w_2} \\ \frac{\partial V_d}{\partial w_1} & \frac{\partial V_d}{\partial w_2} \end{vmatrix} = 0\,. \tag{5.6}$$

If we neglect size effects and use the bulk value for the copper resistivity, (5.6) using (5.2) and (5.4), will yield

$$\begin{vmatrix} \frac{1}{2mF} & \frac{1}{p_x} \\ \frac{2I_0\rho nF}{tw_1^2} l_1(l_1+1) & \frac{4I_0\rho mF}{tw_2^2} l_1 l_2(l_2+1) \end{vmatrix} = 0\,. \tag{5.7}$$

or simply

$$\frac{\mathbf{w}_2}{\mathbf{w}_1} = 2\sqrt{l_2(l_2+1)} \approx 2l_2 = \frac{p_y}{2mF}\,, \tag{5.8a}$$

$$\frac{\mathbf{v}_{d2}}{\mathbf{v}_{d1}} = \frac{p_y}{p_x} = a_r\,, \tag{5.8b}$$

where $a_r$ is the ratio of power via pitch in the $y$ direction to its pitch in the $x$ direction and $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{v}_{d1}$, and $\mathbf{v}_{d2}$ represents the optimal value for $w_1$, $w_2$, $v_{d1}$, and $v_{d2}$, respectively. Replacing (5.8) and (5.2) in the constraint given in (5.5) will result in

$$\mathbf{w}_1 = \frac{I_0 \rho}{4nFtV_{IR}} p_x^2 (1 + a_r). \tag{5.9}$$

The optimal area for a power distribution network in the first two metal levels will be

$$\frac{\mathbf{A}_{power}}{A_{chip}} = \frac{\mathbf{w}_1}{2mF}(1 + a_r). \tag{5.10}$$

### 5.3.2 Considering Size Effects

Considering that the resistivity of copper wires in M1 and M2 depends on the size of the wires, (5.6) will result in

$$\frac{\mathbf{w}_2}{\mathbf{w}_1} = 2l_2 \sqrt{\frac{\left(\partial\rho/\partial w\right)\big|_{w_2} w_2 - \rho(w_2)}{\left(\partial\rho/\partial w\right)\big|_{w_1} w_1 - \rho(w_1)}}. \tag{5.11}$$

Surface scattering and grain boundary scattering are considered the two major mechanisms for size effects. The Fuchs-Sondheimer model presented in [14] is used to model the surface scattering. In this model, an empirical parameter $p$, called specularity parameter, is used to describe the fraction of electrons scattered specularly at the surface. The resistivity of a wire with width $w$ and thickness $t$ can be written as

$$\frac{\rho}{\rho_b} = 1 + \frac{3}{8}(1 - p)\lambda \frac{w + t}{wt}, \tag{5.12}$$

where $\rho_b$ is the bulk resistivity and $\lambda$ is the mean free path of the elections. The grain boundary scattering is modeled as the Mayas-Shatzkes method [65]. In this model, another empirical parameter $R$, called the reflectivity coefficient, describes the probability of electrons getting reflected at the grain boundaries. Based on this model, resistivity can be written as

$$\frac{\rho}{\rho_b} = \frac{1}{3}\left(\tfrac{1}{3} + \tfrac{\alpha}{2} + \alpha^2 - \alpha^3 \ln(1+\tfrac{1}{\alpha})\right)^{-1} \quad , \qquad (5.13)$$
$$as \qquad \alpha = \frac{R}{1-R}\frac{\lambda}{D}$$

where $D$, the average grain diameter, is assumed to be equal to the wire width or thickness, whichever is smaller. In this case, the wire resistivity can be approximated with an error less than 15% as

$$\rho(w) = \rho_0(1 + \beta\frac{w+t}{wt}) , \qquad (5.14)$$

where $\beta$ is a function of $p$, the specularity parameter, and $R$, the reflection coefficient at grain boundaries. Although the value of $\beta$ needs to be calculated numerically, this approximation enables us to derive closed-form expressions for optimal values of wire widths for the power lines in M1 and M2 where the final expressions are independent of $\beta$. Using the approximation given in (5.14), the expression in (5.11) can be expanded as

$$\sqrt{\frac{\left(\partial\rho/\partial w\right)\big|_{w_2} w_2 - \rho(w_2)}{\left(\partial\rho/\partial w\right)\big|_{w_1} w_1 - \rho(w_1)}} = 1 + \beta(\frac{1}{w_2} - \frac{1}{w_1}) + h.o.t. , \qquad (5.15)$$

where *h.o.t.* represents higher-order terms. Using (5.14), $\rho(w_2)/\rho(w_1)$ also will be expanded as

$$\frac{\rho(w_2)}{\rho(w_1)} = 1 + \beta(\frac{1}{w_2} - \frac{1}{w_1}) + h.o.t. \qquad (5.16)$$

Therefore, (5.11) will be simplified as

$$\frac{\mathbf{w}_2}{\mathbf{w}_1} \approx 2l_2\frac{\rho(w_2)}{\rho(w_1)} = \frac{p_y}{2mF}\frac{\rho(w_2)}{\rho(w_1)} . \qquad (5.17)$$

Hence,

$$\frac{\mathbf{v}_{d2}}{\mathbf{v}_{d1}} = \frac{p_y}{p_x} = a_r . \qquad (5.18)$$

Interestingly, (5.18) is the same as (5.8b); the optimal values of $w_1$ and $w_2$ that minimize the total area of the power distribution network are such that the ratio of the *IR* drop in M2 to the *IR* drop in M1 is equal to the ratio of the global via pitches in the *y* and *x* directions. This result is independent of the specularity parameter or scattering probability at the grain boundaries. By substituting (5.18) and (5.2) in the constraint given in (5.5), the optimal width for a given IR drop, and the minimum area needed for the local power distribution network, become

$$\mathbf{w}_1 = \frac{I_0 \rho(w_1)}{4nFtV_{IR}} p_x^2 (1 + a_r), \tag{5.19.a}$$

$$\frac{\mathbf{A}_{power}}{A_{chip}} = \frac{\mathbf{w}_1}{2mF} (1 + a_r \frac{\rho(w_2)}{\rho(w_1)}), \tag{5.19.b}$$

which show that the optimal wire width for the power distribution network depends on the pitch of the power vias that connect the local power distribution network to the global power grid. Also, the area needed for the local power distribution network only depends on the width of the power distribution network in the first metal level.

### 5.3.3 Sub-Optimal Design

For most cases, as shown in (5.8) and (5.17), $\mathbf{w}_2$ is much larger than $\mathbf{w}_1$. However, there are practical and process limitations for maximum and minimum wire widths. The wire width cannot be smaller than the minimum feature size, and it should not be larger than the maximum value, $w_{2\text{-}max}$, that the chemical mechanical polishing (CMP) process allows [8]. If $\mathbf{w}_1 < F$ or $\mathbf{w}_2 > w_{2\text{-}max}$, the designers have to use a larger value for $w_1$ and a smaller value for $w_2$ for the same IR-drop budget. The total metal area needed for power distribution in M1 and M2, however, would be larger, as shown in Figure 5.2. In this figure the for different values of $a_r$, the ratio of the power via in the *y* direction to the *x* direction, the normalized area needed for the local power distribution network and the normalized width of the power lines in M2 are plotted versus the normalized width of the

power lines in M1. For example, by using $w_1$ twice its optimal value ($w_1$=2 $\mathbf{w}_1$), $w_2$ will reduce to 60% of its optimal value ($w_2$=0.6 $\mathbf{w}_2$); however the area consumed by the local power distribution network will increase by 30% ($A_{power}$=1.3 $\mathbf{A}_{power}$).

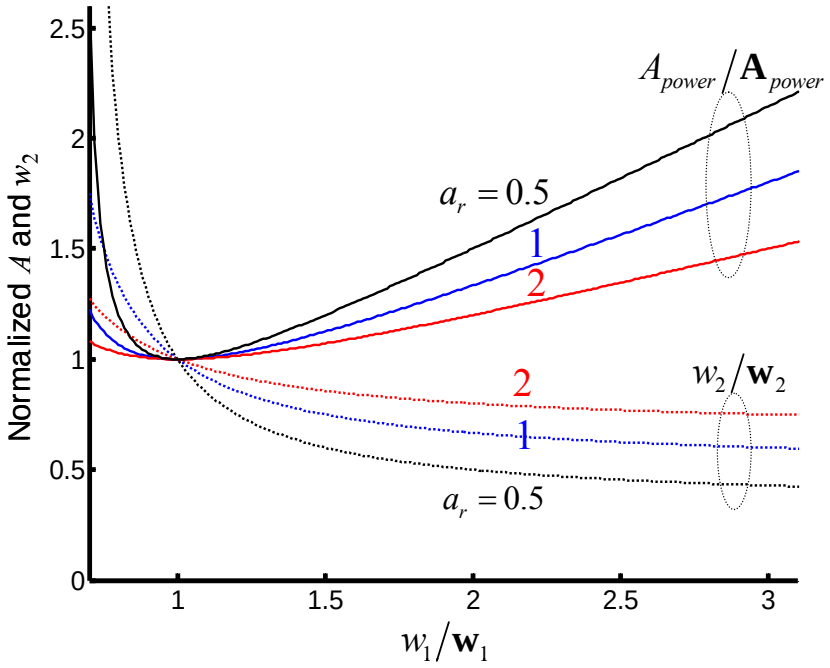



Figure 5.2: Sub-optimal designs: Normalized total metal area in M1 and M2 and the width of power lines in M2 versus normalized width of power lines in M1 for different power via pitch configurations ($a_r$). Total area and wire widths are normalized to their optimal values.

### 5.3.4 Results and Discussion

In this section the optimal design for two cases, with and without size effects, are compared in order to quantify the impact of size effects. Two design scenarios are considered. The first one is to find the optimal values for $w_1$ and $w_2$ (area needed for power distribution) while power via pitches ($p_x$ and $p_y$) are given. The second scenario is to find the pitch of the power vias ($p_x$ and $p_y$) while the area specified for power distribution network

($w_1$ and $w_2$) is fixed. In the first scenario, (5.19.a) shows that for a given IR drop budget, the power wire width is proportional to the resistivity of wires ($\mathbf{w}_1 \propto \rho$ ($w_1$) and $\mathbf{w}_2 \propto \rho$ ($w_2$)). Therefore to compensate, i.e., a four times increase in the resistivity of minimum size wires, $\mathbf{w}_1$ should increase by the same factor. As $w_2$ is larger than $w_1$, the increase in the resistivity of power lines in M2 will be less than four times, so the needed increase in the size of power straps in M2 will be less 4. Hence, totally, more than a three times increase in the wiring area for power lines in the first two metal levels is needed in this case.

In the second scenario, as (5.19.a) implies for fixed $w_1$, the power via pitch should scale down proportionally to the square root of the resistivity of the power wires. Hence, to compensate a four times increase in the resistivity of the minimum-size wires, power

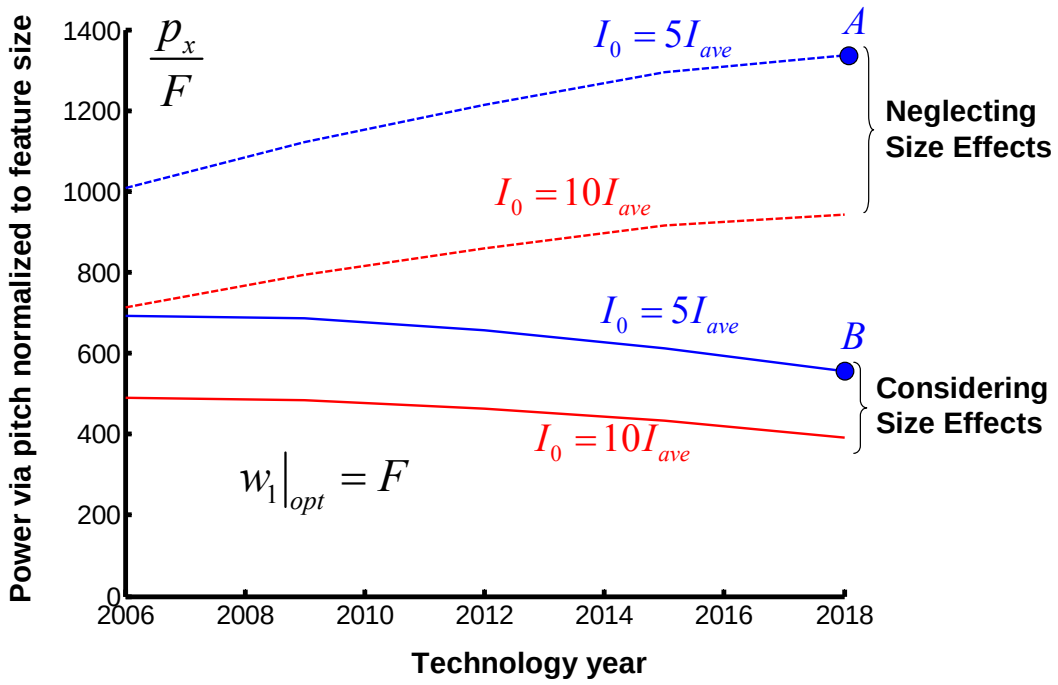



Figure 5.3: Power via pitch normalized to the feature size versus technology year such that the optimal rail width (in M1) is equal to the minimum feature size, $F$. Points A and B are used to show that a 57% reduction in power via pitch is needed to compensate size effects at the year 2018.

via pitches ($p_x$ and $p_y$) should decrease by a factor of 2, or the number of power vias should increase by a factor of 4. Equation (5.19) shows that the optimal value for $w_1$ depends on the $p_x$ and the area dedicated to the power distribution network only depends on $w_1$. Therefore, a $p_x$ that results in a $w_1$ equal to the minimum feature size will optimally consume the minimum possible area in the first two metal levels for the power distribution network. In Figure 5.3, the power via pitch normalized to the feature size is plotted versus technology years such that the optimal value for $w_1$ is equal to the minimum feature size. The impact of chip power consumption on the local power distribution network is shown by means of $k$, the ratio of the maximum to the average gate current. Two values of 5 and 10 are considered for $k$.

Two cases are considered, CASE I, a hypothetical case with no size effect, and CASE II, a case with size effects as described in Section 5.3.2. Since $\mathbf{w}_1$ is chosen to be

Table 5.1: Required power wire widths in M1 and M2 at the end of the ITRS (14-nm node) for the total IR-drop in M1 and M2 of 2% of supply voltage for various global via pitches ($l_1$ and $l_2$). Wire dimensions are normalized to the minimum feature size, and the total metal area, $A$, is normalized to the total chip area. Bold numbers represents the optimal design.

| | $I_0/I_{ave}$=5 | | | | | | $I_0/I_{ave}$=10 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No Size Effects | | | $p$ = 0, $R$ = 0.5 | | | No Size Effects | | | $p$ = 0, $R$ = 0.5 | | |
| | $w_1$/F | $w_2$/F | $A/A_{chip}$ | $w_1$/F | $w_2$/F | $A/A_{chip}$ | $w_1$/F | $w_2$/F | $A/A_{chip}$ | $w_1$/F | $w_2$/F | $A/A_{chip}$ |
| $l_1$ = 40<br>$l_2$ = 10 | 1 | 8.2 | 2.2% | **2.6** | **44.7** | **5.5%** | **1.3** | **22.4** | **3.3%** | **5** | **83** | **9.6%** |
| | | | | 5.3 | 27.3 | 7% | | | | 10 | 54.6 | 12.5% |
| | | | | | | | | | | 15 | 50 | 16.5% |
| $l_1$ = 40<br>$l_2$ = 8 | 1 | 5.7 | 2% | **2.4** | **32.5** | **4.8%** | **1.2** | **16.9** | **2.9%** | **4.6** | **64.5** | **8.4%** |
| | | | | 4.8 | 19.8 | 6.3% | | | | 5.6 | 50 | 8.8% |
| | | | | | | | | | | 9.1 | 39.7 | 11.1% |
| $l_1$ = 50<br>$l_2$ = 8 | 1 | 9.7 | 2.2% | **3.3** | **46** | **5.8%** | **1.7** | **24** | **3.4%** | **6.3** | **91.5** | **10.3%** |
| | | | | 6.6 | 35.7 | 7.9% | | | | 12.7 | 51.6 | 14.3% |
| | | | | | | | | | | 13.7 | 50 | 15.2% |

equal to the minimum feature size, the area dedicated to power wires in the first two metal levels is fixed (around 11%). Points A and B in Figure 5.3 show that for a given constant area dedicated to the local power distribution network, the power via pitch at the end of the ITRS must decrease by 57% to compensate the increase in resistivity by size effects. It also indicates that, along technology nodes, to minimize the area consumption in the first two metal levels by the power distribution network, the power via pitch normalized to the feature size should remain almost constant. Therefore at the same rate as the technology advances, the power via pitch scales down and the number of power vias increases. This will increase the via blockage, especially for the top metal levels, which is the subject of the next section.

Table 5.1 summarizes the results of the local power distribution network design for a chip at the end of the ITRS (14-nm node) while the IR drop is 2% of the supply voltage. The bold numbers represent the optimal designs, and the rest are sub-optimal designs. In different rows and columns, the impact of supply current per gate, size effects, power via pitch, and wire width (sub-optimal design) on the area consumed by local power distribution network is shown.

## 5.4 Area Optimization Considering Via Blockage

It has been shown that for a fixed area dedicated to the local power distribution network, the number of power vias should increase as technology advances. In this way, the burden of the power network is imposed on the metal levels between the local and global power distribution networks. Hence, to find the optimal value for both widths of the local power lines and the power via pitch, the area consumed at the first two metal levels and the area consumed by the power via blockage should be considered at the same time.

### 5.4.1 Power Via Blockage

As shown in Figure 5.1, the local power distribution network is connected to the global power grids through power vias. As discussed in [78], the implicit impact of a via is more than just its physical footprint. For a given metal level, the via blockage factor, which is defined as the ratio of the unused wiring area due to vias, $A_v$, to the chip area, $A_{chip}$, can be expressed as follows:

$$B_v = \frac{A_v}{A_{chip}} = \frac{2W + s\lambda}{\sqrt{A_{chip}/N_v}} \tag{5.20}$$

where $W$ is the wiring width on a given level and assumed to be half of the wiring pitch, $s$ is the via covering factor (~ 0.3), and $\lambda$ is the layout rule unit, which is usually equal to half of the minimum feature size [78]. Assuming the geometry of power via as shown in Figure 5.1, the number of power vias, $N_{v\text{-}power}$, can be written as $N_{v\text{-}power} = A_{chip}/(p_y.p_x)$. Therefore, the power via blockage factor for a pair of metal levels can be written as

$$B_{v-power} = \frac{A_{v-power}}{A_{chip}} = 2\frac{P + s\lambda}{\sqrt{p_x p_y}} \tag{5.21}$$

where $P$ is the wiring pitch. A 57% decrease in power via pitch at the end of the roadmap will result in a 135% increase in the power via blockage factor for all metal levels between M2 and the top-most levels with global power grids. Although the power via blockage is negligible for the local levels, 0.62% for $P$=2$F$, it will consume 10% of the global metal levels if their wire width is, for instance, 20 times the minimum feature size ($P$=40$F$). Assuming that the power vias connecting the local power distribution network to the global power grid, which is located in the $n$-th pair of metal level, the total power via blockage will be

$$B_{v-power} = \frac{A_{v-power}}{A_{chip}} = 2\frac{\sum_{2}^{n-1} P_i + s\lambda(n-2)}{\sqrt{p_x p_y}} \tag{5.22}$$

where $P_i$ is the wiring pitch in the $i$-th pair of the metal level. Therefore, to evaluate the power via blockage, the wiring pitches in the first $n$ metal pairs are needed. In the next section the n-tier methodology is used to determine power via blockage.

### 5.4.2 N-tier Interconnect Network Design

The n-tier methodology optimally designs the wire pitch of each orthogonal pair of metal levels such that the longest interconnect in each level meets the given time constraints [54]. For the 14-nm node (year 2020), the n-tier architecture is designed once, assuming there are no size effects (CASE I), and then the n-tier architecture is redesigned considering size effects (CASE II). The results are shown in the Table 5.2. The results of this section are used to evaluate the power via blockage as given in (5.22).

Table 5.2: Results of n-tier design methodology neglecting (CASE I) and considering (CASE II) size effects for year 2020 (14-nm node). $L_{max}$ is normalized to the gate pitch (288 nm).

| | *CASE I* | | *CASE II* | |
|---|---|---|---|---|
| *# of pairs (Np)* | 6.17 | | 7.3 | |
| | *Pitch* (*nm*) | $L_{max}$ | *Pitch* (*nm*) | $L_{max}$ |
| *pair #1* | 28 | 101 | 28 | 101 |
| *pair #2* | 47 | 862 | 58 | 575 |
| *pair #3* | 116 | 2116 | 114 | 1389 |
| *pair #4* | 207 | 3782 | 184 | 2507 |
| *pair #5* | 333 | 6075 | 270 | 3979 |
| *pair #6* | 574 | 10448 | 387 | 6015 |
| *pair #7* | 1007 | 18400 | 594 | 9712 |
| *pair #8* | - | - | 1075 | 18400 |

### 5.4.3 Optimization

In this section, the optimal value for the power via pitch that minimizes the total area consumed by the power network (considering both local power distribution network and the power via blockage) is studied. By replacing (5.19a) in (5.19b), the optimal area needed for the local power distribution network as a function of power via pitch will be written as

$$\frac{\mathbf{A}_{power}}{A_{chip}} = \frac{p_x^2}{2S_0}$$
$$where \;\; S_0 = \frac{2nFtV_{IR}}{I_0\rho_1(1+a_r)(1+a_r\rho_2/\rho_1)} \tag{5.23}$$

Assuming $a_r=p_x/p_y$, using (5.22), the area blocked by the power via as a function of power via pitch will be

$$\frac{A_{v-power}}{A_{chip}} = \frac{S_v}{p_x}$$
$$where \;\; S_v = \frac{2}{\sqrt{a_r}}[\sum_{2}^{n-1} P_i + s\lambda(n-2)] \tag{5.24}$$

The total area dedicated to the power distribution network, consisting of the local power distribution network and the area blockage due to the power vias, will be written as

$$\frac{A_{tot-power}}{A_{chip}} = \frac{\mathbf{A}_{power} + A_{v-power}}{A_{chip}} = \frac{p_x^2}{2S_0} + \frac{S_v}{p_x} \tag{5.25}$$

which is minimized at

$$p_{x-opt} = \sqrt[3]{S_0 S_v} \tag{5.26}$$

There is an upper and lower limit for the $p_x$. As shown in (5.19.a), $\mathbf{w}_1$ (therefore $\mathbf{w}_2$) depends on $p_x$ ; therefore, lithography limitations ($\mathbf{w}_{1\text{-min}}=F$) put a lower bound for the $p_x$ and process limitations ($w_{2\text{-}max}$) put an upper bound for $p_x$ .

### 5.4.4 Results and Discussion

For a chip in the year 2020, using a multi-level interconnect network as designed in Table 5.2, $\mathbf{A}_{\text{power}}$, $A_{v\text{-}power}$, and $A_{tot\text{-}power}$ normalized to the chip area are plotted versus $p_x$ normalized to the minimum feature size in Figure 5.4 ($a_r$ is assumed to be 1). It shows that for small $p_x$, the number of power vias will be larger; therefore the $A_{v\text{-}power}$ will be larger than $\mathbf{A}_{\text{power}}$ and vice versa for large $p_x$. The total area is minimized for $p_x$=690$F$. This point corresponds to $\mathbf{w}_1$=2.7$F$. The total area is also plotted versus power via pitch neglecting size effects. In this case, the total area is minimized for $p_x$=1060$F$ ($\mathbf{w}_1$=1.5$F$). The plot

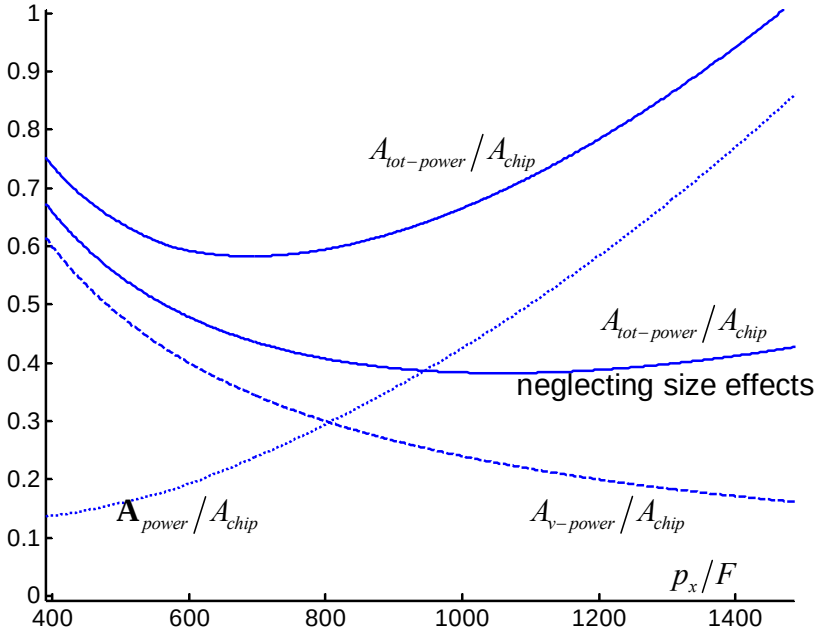



Figure 5.4: Area needed for the local power distribution network (dotted line), via blockage (dashed line) and the total area (solid line) normalized to the chip area are potted versus power via pitch in the $x$ direction normalized to the minimum feature size. This plot is for a case study in year 2020; size effects are considered and $a_r$=1. The second solid line is the total area normalized to the chip area for the case when size effects are not considered.

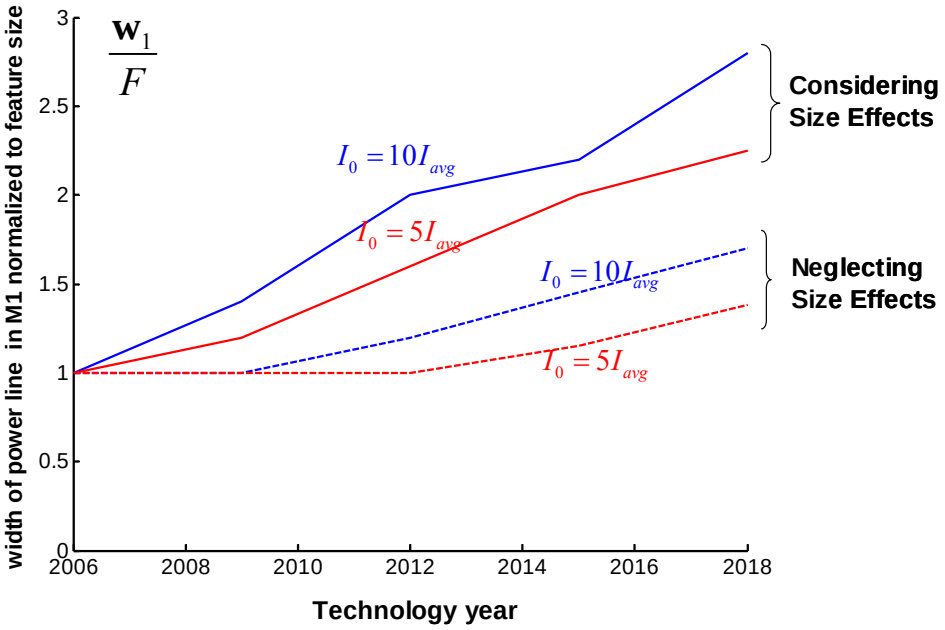
Table 5.3: Results of Local power distribution network design for 14-nm node (year 2020) for different power via ratios in $y$ and $x$ direction ($a_r$). $w_1$, $w_2$, and $p_x$ are normalized to the minimum feature size ($F$=14 nm) and total area needed for the local power distribution ($A_{total}$) is normalized to the chip area ($A_{chip}$)

| | $w_1$ | $w_2$ | $p_x$ | $A_{total}$ |
|---|---|---|---|---|
| 0.5 | 3.6 | 34 | 940 | 60% |
| 0.7 | 3.2 | 34 | 815 | 58.8% |
| 1 | 2.7 | 34 | 690 | 58% |
| 1.5 | 2.3 | 33 | 555 | 59% |
| 2 | 2 | 34 | 470 | 60% |

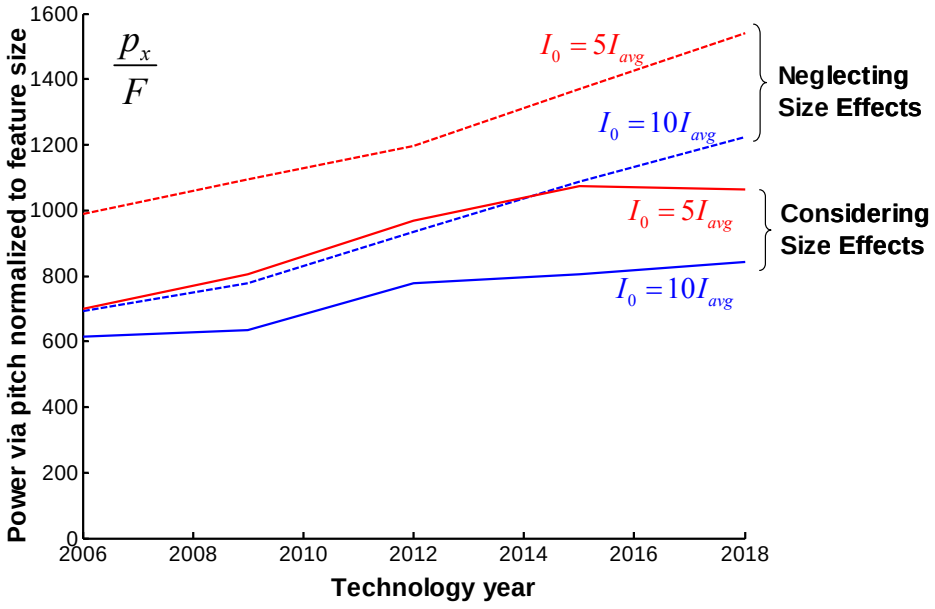
shows that in the case that neglects size effects by using optimal design, 38% of the chip area will be consumed for the local power distribution network (considering via blockage). If we use the same via pitch for the case considering size effect, 70% of the chip area will be consumed by the local power distribution network. However, by redesigning the power distribution network (using $p_x$=690$F$), the area consumption will be reduced to 58%. These results show that it is important to consider size effects in the design process of the local power distribution network. For a fixed power via pitch, 84% more copper in the first two metal levels is needed to compensate size effects; however, by redesigning the power via pitch (considering size effects in the design), size effects could be addressed by utilizing only 52% more area.

The results of the optimal design of a case study at the 14-nm node for different $a_r$ (the ratio of power via pitch in the $y$ direction to its pitch in the $x$ direction) are summarized in Table 5.3. Although $A_{total}$ shows small dependency on $a_r$, $a_r$=1 offers the least area consumption by the local power distribution network. Power via pitch ($p_x$) that minimizes $A_{power}$ for $\mathbf{w}_1$=$F$ versus technology years was shown in Figure 5.3. Considering power via blockage, $\mathbf{w}_1$=$F$ will not be the optimal design. This is shown in Figure 5.5.a. The optimal width of the power rail in M1 that minimizes $A_{total}$ is plotted versus technology years, considering and neglecting size effects. The impact of chip power

consumption is shown by means of a $k$ factor of 5 or 10. It shows that for the year 2006, the impact of power via blockage is negligible; therefore $\mathbf{w}_1=F$ for all cases, while at the end of the roadmap, $\mathbf{w}_1$ should be up to three times the minimum feature size. Considering $\mathbf{w}_1$ given in Figure 5.5.a, power via pitch normalized to the feature size is plotted versus technology year in Figure 5.5.b. The power via pitch at year 2006 is the same as what was shown in Figure 5.3, but as the technology advances, considering power via blockage, the optimal $p_x$ is larger than optimal $p_x$ neglecting power via blockage.




(a)




(b)

Figure 5.5: (5.5-a) Width of the optimal power line in M1 normalized to the feature size vs. technology year. (5.5-b) Power via pitch normalized to the feature size versus technology year such that the optimal rail width (in M1) is given in (5.5-a) minimizing the total area consumed both in M1 and M2 and power via blockage.

## 5.5 Conclusions

An optimization methodology for local power distribution networks is presented that, for a given *IR* drop budget, minimizes the total area needed for power lines. The result shows that the optimal wire widths are such that the ratio of *IR* drop in M2 to the *IR* drop in M1 is equal to the ratio of the global via pitches in the *y* and *x* directions. The case study for the 14-nm node shows that to compensate up to four times increase in the resistivity of minimum size wires, either a three times increase in wiring area for power lines in the first two metal levels would be needed or a two times decrease in the via pitch would be necessary to address the impact of size effects.

It has been shown that for a fixed area dedicated to the local power distribution network, the pitch of power vias between the local and global power distribution network should decrease as technology advances. However, decreasing the power via pitch will increase the power via blockage factor for all metal levels between M2 and the top-most levels with global power grids. The results shows that for 2020, the optimal design that minimizes both the area needed for the local power distribution and power via blockage factor occurs when power via pitches in the *x* and *y* directions are equal and widths of the power lines in M1 are three times the minimum size wires in M1.

# CHAPTER 6

# Conclusions and Future Work

In this chapter, several important extensions to this thesis for pursuing further work in this area are briefly explored. Finally, the main contributions and conclusions of our results are presented.

## 6.1 Future Work

Based on this work, several questions can be raised regarding the operation of interconnects at high frequency and low dimensions. These questions can open different avenues for future investigations.

### 6.1.1 Size Effects and ASE Modeling

Size effects and the anomalous skin effect have been the subject of many previous studies. In all of these studies based on Matthiessen's rule, the resistivity is considered as

$$\rho(T) = \rho_{e-p}(T) + \rho_{defects} \tag{6.1}$$

where $\rho_{e-p}$ is the resistivity due to electron and phonon scattering and $\rho_{defects}$ is the resistivity due to defect scattering that is independent of temperature and includes surface and sidewall (roughness) scattering, grain boundary scattering, and other structural defects and impurities. It is usually assumed that the scattering mechanisms are independent of each other. Hence, as shown in Figure 4.2, the total increase in resistivity is considered to be the summation of the increase in resistivity due to surface scattering and grain boundary scattering. Assuming that the size of the grains is equal to the wire width, this assumption is valid, but recent studies of the distribution of the grain-sizes show that size of many of the grains could be smaller than the wire dimensions [79]. If so, many of the electrons located inside in the smaller grain have no chance to be scattered from the

surfaces. Hence, Matthiessen's rule is not applicable and surface scattering and grain boundary scattering should be considered simultaneously.

The theory of ASE is developed for metals with no grain boundaries. However, copper wires in GSI applications are polycrystalline. Grain boundaries disrupt the motion of electrons. The normal or anomalous skin effect has not yet been studied for a polycrystalline copper wire.

### 6.1.2 Interconnects at Low Temperature

It has been known since the beginning of CMOS electronics that the chip performance enhances by lowering the temperature. Among the advantages of sub-ambient cooling are faster switching times, increased circuit speed, and a reduction in the thermally induced failures of the devices [80]. A metric that characterizes one of the fundamental limits for the gigascale integration of digital electronics is the signal energy transfer that occurs during the binary transition. It can be written as $E_{\min} = kT \ln 2$ , where $T$ is the absolute temperature, and $k$ is Boltzmann's constant [81]. Therefore, the only method to improve this fundamental limit is to operate at lower temperatures. While, we are still far from this fundamental limit in today's CMOS technology, we all know that we are "power limited." It has been shown that for every 10°C reduction in the transistor's temperature, chip performance improves from 1 to 3% [80, 82].

As shown in Figure 1.2 and argued in Section 2.5 of this thesis, the second corner frequency for ASE (the frequency above which the resistivity of copper wires depends on the frequency to the power of 2/3) is a function of temperature and will be less at lower temperatures. Hence, in this regime the surface impedance of a copper wire can be written as

$$Z_{surf} = \frac{1}{4}\left(\frac{12}{\pi^2}\right)^{1/6}\left(\frac{\rho}{\lambda}\right)(1+\sqrt{3}j)\left(\frac{\lambda}{\delta}\right)^{4/3} \tag{6.2}$$

where $\rho$ is the resistivity, $\lambda$ is the mean free path, and $\delta$ is the classical skin depth. Therefore, impedance per unit length of the transmission line as previously shown in (3.34) should be written as

$$z(s) = r_{dc} + Ks^{2/3} + l(s)s \tag{6.3}$$

where $l(s) \propto s^{-1/3}$. However, the phasor analysis that we developed in Chapter 3 is capable of modeling a transmission line with impedance per unit length as in (6.3). It will be very interesting to study the step response of such a wire. This analysis could lead to a new bit-rate formula for metallic wires at low temperatures.

### 6.1.3 Multi Level Interconnect Network Design

In Chapter 4, we introduced an optimization methodology that optimized the reverse-scaled multi-level interconnect network to reduce the logic macrocell area, cycle time, power consumption, or number of metal levels. This methodology uses an a priori wiring distribution. The wiring efficiency (the ratio of the area available for signal wires per metal level to the chip area) is assumed to be constant for all metal levels. In Chapter 5, an optimization methodology for the local power distribution network is presented. It has been shown that the area blockage due to the power vias should be considered in the area minimization of a local power distribution network.

Optimization methodologies for mulit-level interconnect network design that considers signal, power, and clock wires could be the next possible step. The local power distribution network designed in Chapter 5 provides an a priori estimation for the power via blockage. The power and signal via blockage reduces the wiring efficiency. In the local levels, the signal via blockage is dominant and the power via blockage is negligible, while at the global levels, the power via blockage is considerable and the signal via blockage is negligible. A multi-level interconnect network should consider clock distribu-

tion wires, power distribution network, and power via blockage in the wiring efficiency of the available wiring area.

## 6.2 Conclusions and Contributions

A new physical model for the anomalous skin effect is developed. Based on this model, effective resistivity of thin wires with a rectangular cross-section as a function of wire dimension and frequency is extracted. A compact formula for the resistivity of copper wires considering the impact of line-edge roughness is presented.

A phasor analysis technique is developed to model a transmission line with general series impedance and parallel admittance per unit length. It is shown that the distributed model (quasi-TEM approximation) is valid for on-chip interconnects until the end of the roadmap for semiconductors. The delay analysis shows that for different wire sizes, the impact of ASE on delay calculations can be neglected.

A compact model for the bit-rate limit of a transmission line is presented. This new model is capable of considering skin effect, surface scattering, and dc resistance simultaneously. The low-loss approximation that is used in the previous models is valid only when the bit-rate limit is larger than the reciprocal time-of-flight. In contrast to the previous models, it is shown that the bit-rate limit is not scale-invariant. It is shown that ASE should be considered for on-chip interconnects with bit-rate limits over 20 GHz. This model is used to study the trade-off between the electrical and mechanical compliance of the dual-mode polymer pins.

It is shown that for high-performance applications, modest changes in the design of a multi-level interconnect network can address the increases in copper resistivity resulting from surface and grain boundary scatterings such that the longest interconnect in each metal level meets the specified time constraint. The main reason for the modest changes is that interconnects that are affected most by size effects are so short that their latencies are not important. Slight changes in wiring pitches allow for transferring the critical ones

that are hurt by size effects to higher metal levels with larger wiring pitches such that they meet the specified timing constraint. For a case study chip implemented at the 18-nm node, it is shown that the changes in the number of metal levels will be less than 7%.

It is shown that for low-cost applications, very few wiring pitches are normally used and the number of metal levels needed to compensate the impact of size effects on the average *rc* delay of copper interconnect is drastically high. For a case study chip implemented at 14 nm with a two-tier interconnect network, it is shown that a 60% increase in the number of metal levels is needed to address a five times increase in the resistivity of minimum-size wires.

An optimization methodology for local power distribution networks is presented that, for a given IR drop budget, minimizes the total area needed for power lines. The result shows that the optimal wire widths are such that the ratio of the IR drop in Metal 2 to the IR drop in Metal 1 is equal to the ratio of the global via pitches in the *y* and *x* directions. It has been shown that for a fixed area dedicated to the local power distribution network, the pitch of power vias between the local and global power distribution network should decrease as technology advances. However, decreasing the power via pitch will increase the power via blockage factor for all metal levels between M2 and the top-most levels with global power grids. The results shows that for 2020, the optimal design that minimizes both the area needed for the local power distribution and power via blockage factor occurs when power via pitches in the x and y directions are equal and widths of the power lines in M1 are three times the minimum size wires in M1.

# REFERENCES


1. Moore, Gorden E., *Cramming more components onto integrated circuits*. Electronics Magazine, Vol 38, No. 8, April 1965. [Available from: http://www.intel.com/technology/mooreslaw/, June 2008]

2. Jeffrey A. Davis and James D. Meindl, *Interconnect technology and design for gigascale integration*. Boston: Kluwer Academic Publishers, 2003.

3. Davis, J.A.; Venkatesan, R.; Kaloyeros, A.; Beylansky, M.; Souri, S.J.; Banerjee, K.; Saraswat, K.C.; Rahman, A.; Reif, R.; Meindl, J.D., *Interconnect limits on gigascale integration (GSI) in the 21st century.* Proceedings of the IEEE, vol. 89, pp. 305 - 324, Mar 2001.

4. Venkatesan, R.; Davis, J.A.; Bowman, K.A.; Meindl, J.D., *Optimal n-tier multilevel interconnect architectures for gigascale integration (GSI).* IEEE Trans. VLSI Syst., vol. 9, pp. 899-912, Dec 2001.

5. Edelstein D.C.; Sai-Halasz G.A.; Mii Y.-J, *VLSI on-chip interconnection performance simulations and measurements.* IBM Journal of Research and Development, vol.39, Issue 4, pp. 383-401, 1995.

6. E.V.Barnat, D.Nagakura, P.-I.Wang, and T.-M.Lu, *Real time resistivity measurements during sputter deposition of ultrathin copper films.* J. Appl. Phys., vol. 91, Issue 3, pp. 1667-1672, 2002.

7. *International Technology Roadmap for Semiconductors*. 2006 [cited; Available from: http://public.itrs.net/ , June 2008].

8. Geraldine Cogin; Schwartz, K. V; Srikrishnan, Arthur Bross, *Handbook of Semiconductor Interconnection Technology*. CRC Press. 2006

9. Powell, R. *Tutorial on Cu Resistivity*. in *Copper Resistivity Workshop*. 2005.

10. Pippard, A.B., *The Surface Impedance of Superconductors and Normal Metals at High Frequencies. II. The Anomalous Skin Effect in Normal Metals.* Proc. Roy. Soc. London A, 1947. vol.191, Issue 1026, p. 385-399, Nov. 1947.

11. David Jiles, *Introduction to the Electronic Properties of Materials*. CRC Press, 2001.

12. Chambers, R.G., *The Anomalous Skin Effect.* Proc. Roy. Soc. London A, vol. 215, Issue 1123, pp. 481-497, Dec. 1952

13. Dingle, R.B., *The Electrical Conductivity of Thin Wires*. Proc. Roy. Soc. London A, vol.201, Issue 1067, pp. 545-560, May 1950.

14. Sondheimer, E.H., *Mean Free Path of Electrons in Metals.* Advan. Phys., vol. 1, Issue 1, pp. 1-42, Jan 1950.

15. Brews, J.R., *Transmission line models for lossy waveguide interconnections in VLSI.* IEEE Transactions on Electron Devices, vol. 33, pp. 1356-1365, Sep 1986.

16. D. A. B. Miller and H. M. Ozaktas, *Limit to the Bit-Rate Capacity of Electrical Interconnects from the Aspect Ratio of the System Architecture.* Journal of Parallel and Distributed Computing, vol. 41, pp. 42-52, 1997.

17. Svensson, C., *Electrical Interconnects Revitalized.* IEEE transaction on VLSI, vol. 10, Issue 6**, p**p. 777-788, Dec 2002.

18. A. Naeemi; Jianping Xu; A.V. Muleapos;; T.K. Gaylord; J.D. Meindl, *Optical and electrical interconnect partition length based on chip-to-chip bandwidth maximization,*. Photonics Technology Letters, vol. 16, Issue 4, pp.1221 - 1223, April 2004.

19. Sarvari, R.; Naeemi, A.; Meindl, J.D., *General Compact Model for Bit-Rate Limit of Electrical Interconnects Considering DC Resistance, Skin Effect and Surface Scattering*. in *Proceedings of the IEEE International Interconnect Technology Conference*. pp. 163- 165, June 2004.

20. Davis, J.A.; De, V.K.; Meindl, J.D., *A Stochastic Wire-Length Distribution for Gigascale Integration (GSI)-Part I.* IEEE Trans. Electron Devices, vol. 45, Issue 3, pp. 580-597, Mar 1998.

21. Dirk Stroobandt, *A Priori Wire Length Estimates for Digital Design*. Springer, 2001.

22. Jari Nurmi, *Interconnect-Centric Design for Advanced Soc and Noc*. Springer, 2004.

23. Yamada, T.; Nakata, Y.; Hasegawa, J.; Amano, N.; Shibayama, A.; Sasago, M.; Matsuo, N.; Yabu, T.; Matsumoto, S.; Okada, S.; Inoue, M., *A 64-Mb DRAM with meshed power line.* IEEE Journal of Solid-State Circuits, vol. 26, Issue 11,pp. 1506-1510, Nov 1991.

24. Andrey V. Mezhiba and Eby G. Friedman ;, *Power Distribution Networks in High Speed Integrated Circuits*. Springer, 2003.

25. Qing K. Zhu, *Power Distribution Network Design for VLSI*. 2004: Wiley-Interscience, 2004.

26. Schindler, G., *Conductor Resistivity Behavior in Deep Sub-100nm Semiconductor Products*. in *Sematech workshop on Cu resistivity*, 2005.

27. Tellier,C. R. and Tosser, A. J., *Size Effects in Thin Films*. Elsevier, 1982.

28. Durkan, C. and Welland, M.E., *Size effects in the electrical resistivity of polycrystalline nanowires.* Physical Rev. B, vol. 61, Issue 20, pp. 14215–14218, 2000.

29. Sarvari, R.; Meindl, J.D., On the Study of Anomalous Skin Effect for GSI Interconnections. in *Proceedings of IITC*, pp. 42 - 44, June 2003.

30. Naeemi, A.; Sarvari, R.; Meindl, J.D., *Performance comparison between carbon nanotube and copper interconnects for gigascale integration (GSI)*. IEEE Electron Devices Letters, vol. 26, Issue 2, pp. 84 - 86, Feb 2005.

31. N.W. Ashcroft and N.D. Mermin, *Solid state physics*. Philadelphia: Saunders College. 1976.

32. Uichiro Mizutani, *Introduction to the electron theory of metals*. New York: Cambridge University Press. 2001.

33. J. M. Ziman., *Electrons and Phonons*. Oxford Unversity Press, 1962.

34. R.D. Barnard, *Thermoelectricity in Metal and Alloys*. London: Taylor & Francis, 1972.

35. Frank J. Blatt, *Physics of electronic conduction in solids*. New York: Mcgraw-hill. 1968.

36. Y. Namba, *Resistivity and temperature coefficient of thin metal films with rough surface.* Jpn. J. Appl. Phys., vol. 9, pp. 1326–1329, 1970.

37. Lopez, G. Murali, R. Sarvari, R. Bowman, K. Davis, J. Meindl, J., *The Impact of Size Effects and Copper Interconnect Process Variations on the Maximum Critical Path Delay of Single and Multi-Core Microprocessors*, in *Proceedings of IITC*. pp. 40-42, June 2007.

38. Steinhögl, W.; Schindler, G.; Steinlesberger, G.; Traving, *Impact of line edge roughness on the resistivity of nanometerscale interconnects*, in *Microelectronic Engineering European Workshop on Materials for Advanced Metallization 2004,*. pp. 126-30, 2004.

39. Harper, C.A., *Electronic Packaging & Interconnection Handbook, Second Ed.* New York,: McGraw-Hill Inc, 1997.

40. Bakoglu, H.B., *Circuits, Interconnections, and Packaging for Vlsi*. New York: Addison-Wesley Publishing Co., 1990.

41. T. C. Edwards and M. B. Steer, *Foundations of Interconnect and Microstrip Design*. 3rd ed. Wiley, 2001.

42. Hugo Veenstra and John R. Long, *Circuit and Interconnect Design for High Bit-rate Applications*. Springer, 2008.

43. Mustafa Celik, Larry Pileggi, and Altan Odabasioglu, *Ic Interconnect Analysis*. Kluwer Academic Pub, 2002.

44. Collin, R.E., *Foundations for Microwave Engineering,* Wiley-IEEE Press, 2002.

45. Grivet, P., *Physics of Transmission Lines at High and Very High Frequencies*. Academic Press, 1970.

46. Simon Ramo, John R. Whinnery, Theodore Van Duzer, *Fields and Waves in Communication Electronics*. 3rd Edition ed., Wiley, 1994.

47. Richard E. Matick, *Transmission Lines and Communication Networks : An Introduction to Transmission Lines, High-frequency and High-speed Pulse Characteristics and Applications*.Wiley-IEEE Press, 1995.

48. Deutsch, A.; Coteus, P.W.; Kopcsay, G.V.; Smith, H.H.; Surovic, C.W.; Krauter, B.L.; Edelstein, D.C.; Restle, P.L., *On-chip wiring design challenges for gigahertz operation.* Proceedings of the IEEE, vol. 89, Issue 4, pp. 529 - 555, Apr 2001.

49. Giovanni Miano , *Transmission Lines and Lumped Circuits: Fundamentals and Applications*. Academic Press, 2001.

50. Collin, R.E., *Foundations for Microwave Engineering,* Wiley-IEEE Press, 2002.

51. Richard E. Matick, *Transmission Lines and Communication Networks : An Introduction to Transmission Lines, High-frequency and High-speed Pulse Characteristics and Applications*.Wiley-IEEE Press, 1995.

52. Horno, M.; Mesa, F.L.; Medina, F.; Marques, R., *Quasi-TEM analysis of multilayered, multiconductor coplanarstructures with dielectric and magnetic anisotropy including substratelosses.* IEEE Transactions on Microwave Theory and Techniques, vol. 38, Issue 8, pp.1059-1068, Aug 1990.

53. Bakir, M.S.; Bing Dang; Ogunsola, O.O.A.; Sarvari, R.; Meindl, J.D., *Electrical and Optical Chip I/O Interconnections for Gigascale Systems.* IEEE Transactions on Electron Devices, vol. 54, Issue 6, pp. 2426-2437, Sept 2007.

54. Venkatesan, R.; Davis, J.A.; Bowman, K.A.; Meindl, J.D., *Optimal n-tier multilevel interconnect architectures for gigascale integration (GSI).* IEEE Trans. VLSI Syst., vol. 9, Issue 6, pp. 899-912, Dec 2001.

55. P. Murarka and S. W. Hymes, *Copper metallization for ULSI and beyond.* Critical Reviews in Solid State and Materials Sciences, vol. 10, No. 2, pp. 87-124, 1995

56. *International Technology Roadmap for Semiconductors*. 2005 [cited; Available from: http://public.itrs.net/ , June 2008]

57. Rossnagel, S. M and Kuan, T. S.., *Alteration of Cu conductivity in the size effect regime.* J. Vac. Sci. Technol. B., vol. 22, Issue 1, pp. 240-247, 2004.

58. W. Steinhoegl, G. Schindler, G. Steinlesberger, and M. Engelhardt, *Size-dependent Resistivity of Metallic Wires in the Mesoscopic Range.* Phys. Rev. B, 2002. vol. 66, p. 075414, 2004.

59. Steinhögl, W.; Schindler, G.; Steinlesberger, G.; Traving, M.; Engelhardt, M., *Comprehensive study of the resistivity of copper wires with lateral dimensions of 100 nm and smaller.* Journal of Applied Physics, vol. 97, 023706-0237067, 2005.

60. Sungjun Im; Srivastava, N.; Banerjee, K.; Goodson, K.E., *Scaling analysis of multilevel interconnect temperatures for high-performance ICs.* IEEE Transactions on Electron Devices, vol. 52, Issue 12, pp. 2710 - 2719 , 2005.

61. Nguyen, V.H. Christie, P. Heringa, A. Kumar, A. Ng, R. ;. *An analysis of the effect of wire resistance on circuit level performance at the 45-nm technology node*. in *Proceedings of the IEEE International Interconnect Technology Conference*. pp. 191- 193 , June 2005.

62. Naveed A. Sherwani, *Algorithms for VLSI Physical Design Automation*. Third ed., Springer, 1999.

63. Davis, J.A.; De, V.K.; Meindl, J.D. *A Stochastic Wire-Length Distribution for Gigascale Integration (GSI)-Part I.* IEEE Trans. Electron Devices, vol. 45, Issue 3, pp. 580-597, Mar 1998.

64. Zarkesh-Ha, P.; Wright, P. ; Lakshminarayanan, S.; Cheng, C.-C. ; Loh, W.; Lynch, W. . *Backend process optimization for 90 nm high-density ASIC chips*. in *Proceedings of the IEEE International Interconnect Technology Conference*. pp. 123- 125, 2003.

65. A. F. Mayadas, M. Shatzkes, and J. F. Janak, *Electrical-Resistivity Model for Polycrystalline Films: The Case of Arbitrary Reflection at External Surfaces.* Phys. Rev. B, vol. 4, p. 1382 - 1389, 1970.

66. Sarvari, R.; Naeemi, A.; Venkatesan, R.; Meindl, J.D., *Impact of Size Effects on the Resistivity of Copper Wires and Consequently the Design and Performance of Metal Interconnect Networks*. in *Proceedings of the IEEE International Interconnect Technology Conference*. pp. 197- 199, June 2005.

67. Sekar, D.C.; Naeemi, A.; Sarvari, R.; Davis, J.A.; Meindl, J.D., *Intsim: a CAD tool for optimization of multilevel interconnect networks*. in *International Conference on Computer-Aided Design*. pp. 560-567, Nov. 2007.

68. Mikhail Popovich, Andrey V Mezhiba, Eby G Friedman ;, *Power Distribution Networks with On-Chip Decoupling Capacitors*. Springer, 2008.

69. Shen Lin; Chang, N.. *Challenges in Power-Ground Integrity*. in *ICCAD*. 2001, 651 - 654, 2001.

70. Xiaohai Wu; Xianlong Hong; Yici Cai; Zuying Luo; Chung-Kuan Cheng; Jun Gu; Dai, W. , *Area minimization of power distribution network using efficient nonlinear programming techniques.* IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 23, Issue 7 , pp.1086-1094, July 2004.

71. Chowdhury, S.; Breuer, M., *Minimal area design of power/ground nets having graph topologies.* IEEE Transactions on Circuits and Systems, vol. 34, Issue 12, pp. 1441-1451, Dec 1987.

72. Tan, S.X.-D.; Shi, C.-J.R., *Efficient very large scale integration power/ground network sizing based on equivalent circuit modeling.* IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 22, Issue 3, pp. 277-284, 2003

73. Shakeri, K.; Meindl, J.D., *Compact physical IR-drop models for Chip/Package Co-Design of Gigascale Integration.* IEEE Trans. Electron Device, vol. 62, Issue 6, pp/ 1087-1096, June 2005.

74. Gang Huang; Sekar, D.C.; Naeemi, A. ;Shakeri, K.; Meindl, J.D.; *Compact Physical Models for Power Supply Noise and Chip/Package Co-Design of Gigascale Integration*. in *ECTC*. 2007, pp. 1659-1666, 2007.

75. Sarvari, R.; Naeemi, A.; Venkatesan, R.; Meindl, J.D., *Impact of Size Effects on the Resistivity of Copper Wires and Consequently the Design and Performance of Metal Interconnect Networks*. in *Proceedings of the IEEE International Interconnect Technology Conference*. pp. 197- 199, June 2005.

76. Naeemi, A.; barvan, A.; Memdl, J.D.; *On-Chip Interconnect Networks at the End of the Roadmap: Limits and Nanotechnology Opportunities*. in *Proceedings of IITC*, pp. 201-203, 2006.

77. Eble, J.C., *A generic system simulator with novel on-chip cache and throughput models for gigascale integration*., PhD. thesis, Georgia Institute of Technology. 1999.

78. Qiang Chen; Davis, J.A.; Zarkesh-Ha, P.; Meindl, J.D., *A compact physical via blockage model.* IEEE Transactions on VLSI Systems, vol. 8, Issue 6. pp. 689 - 692, DEc 2000.

79. Gardner, D.S., *private communication*. 2008.

80. Kirschman, R.K., *Low-temperature Electronics*. IEEE press, 1986.

81. Meindl, J.D.; Davis, J.A., *The fundamental limit on binary switching energy for terascaleintegration (TSI).* IEEE Journal of Solid-State Circuits, 2000. vol. 35, Issue 10, pp. 1515-1516, Oct 2000.

82. Jaeger, R.C. *Development of Low Temperature CMOS for High Performance Computer Systems,*. in *Proceedings of the IEEE International Conference on Computer Design*. pp. 128-130, 1986.

# *VITA*

Reza received the B.S. and M.S. degrees in electrical engineering from Sharif University, Tehran, Iran, in 1996 and 1998, respectively, and the M.S. and PhD degrees in electrical and computer engineering from Georgia Institute of Technology, Atlanta, in 2003 and 2008, respectively. His primary area of interest is modeling of interconnects with nanometer dimensions, operating at gigahertz frequencies. He enjoys playing soccer, blogging and listening to music, on his spare time.